

RESULT 2  
PCT-US03-19834-2  
; Sequence 2, Application PC/TUS0319834  
; GENERAL INFORMATION:

APPLICANT: STAPLETON, JACK T.  
TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND  
FILE REFERENCE: IOWA:045W0  
CURRENT APPLICATION NUMBER: PCT/US03/19834  
PRIOR FILING DATE: 2003-05-27  
PRIOR APPLICATION NUMBER: 10/445,724  
PRIOR FILING DATE: 2002-06-28  
PRIOR APPLICATION NUMBER: 60/392,158  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 3011  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
PCT-US03-19834-2

Query Match 14.0%; Score 254; DB 1; Length 3011;  
Best Local Similarity 31.8%; Pred. No. 2; Ie-14;  
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY 23 RPAVDPREVLYQEFDEMEAS-HLPYIEGQMLAEQFKOKALAF---ASRQNHVSPTHY 78  
DB 1691 KPAIPDREVLYREFDEMECSQHLPIEQQMMLAEQFKOKALGLQTSRQAEV----- 1745  
QY 79 VPESDAPPAIIPDREVLYREFDEMEAFASR-GNHVSPPHYVESDASQAAPYIEQ---- 133  
DB 1746 ----IAPAV-----QTNWQKLETFWAKHMMNFSGIQYLAGSLTLPGNPAIASLMAF 1793  
QY 134 -AOVIAHQEKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170  
DB 1794 TAAVTSPLTTSQTLFNILIGWVAQDLAAGAAATAFVGAGLGAIGSVGLKVLIDILA 1853  
QY 171 -----IPDREVLYREFDEMEESQHL--PYIEGQML-----AE 201  
DB 1854 GYGAGVAGALVAKKINGEVPSTEDLVNLPALISGALVGVVCAILRRHVGPEGAV 1913  
QY 202 QFKOKALAFASRGNHVAPTHYVESDA 228  
DB 1914 QMNRILAFASRGNHVSPTHYVESDA 1940

RESULT 3  
US-10-370-481-31  
; Sequence 31, Application US/10370481  
; GENERAL INFORMATION:  
; APPLICANT: Hitachi, Ltd.  
; TITLE OF INVENTION: AIP-INTERACTING PROTEINS AND USE THEREOF  
; FILE REFERENCE: HITA.0164  
; CURRENT APPLICATION NUMBER: US/10/370,481  
; PRIOR FILING DATE: 2003-02-24  
; PRIOR APPLICATION NUMBER: US 60/358,369  
; PRIOR FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 60/019,808  
; PRIOR FILING DATE: 2002-09-03  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 31  
; LENGTH: 4574  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-370-481-31

Query Match 6.2%; Score 113.5; DB 6; Length 4574;  
Best Local Similarity 20.1%; Pred. No. 0.13;  
Matches 73; Conservative 65; Mismatches 169; Indels 57; Gaps 10;

QY 14 AEOFKOKAL-----RPAVDPREVLY-----YQEFDEMEASHLPYIEQOMOL 55  
DB 2013 AEEVQKSLAAEEBAARQKALAEVERLKANVEEARLREARAEQESARQLQAEAAOK 2072

QY 56 AEOFKOKALAFASRGNHVSPTHYVESDAPPAIIPDREVLYREFDEME-EAFASR--GNH 112  
DB 2073 RLQAEKKAHAFVQK-----EQLQOOLQOESVL-----DQLGEEAARRAAE 2119  
QY 113 VSPAHYVESDASQAAPYIEQNOVIAHQEKEKVLAF--SRGNHDSPTHYVESDAKPAI 171  
DB 2120 AEEARVQAEAREAAQRQVEEAEERLKQAEQQAQAQAQAALREAEQEAARRQA 2179  
QY 172 PDREVLYREFDEMECSQHLPIEQQMMLAEQFKOKALAFASRGNHVAPTHYVESDAKPA 231  
DB 2180 EQAALROKQADAEEMEKHKFAEQTLROKQAVEDELTLRLQ-----LEETDHQKN 2230  
QY 232 LVPDKEVLYQYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEXVLA 291  
DB 2231 LL-----DEELQRLKAEATBAARQSOVEEELFSVRYQWELSKLARIEAENRA 2280  
QY 232 FASRGNHVSPTHYVESEPOVVVTPDKELIYEAPDE---MEBASKALIEGQMAEMLK 348  
DB 2281 LILRKDNTQRFLOEAEKMKQVAEEAARLSVAQEAARLRLQLAEDLAQQRALAEKMLK 2340  
QY 349 SKIQ 352  
DB 2341 EXMQ 2344

RESULT 4  
US-60-479-073-495  
; Sequence 495, Application US/60479073  
; GENERAL INFORMATION:  
; APPLICANT: De Wilde, Gert Jules Hector  
; APPLICANT: Saunders, Michael John Scott  
; APPLICANT: Logghe, Marc Georges  
; TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the  
; TITLE OF INVENTION: prevention and/or treatment of metabolic diseases and nucle  
; FILE REFERENCE: D00590.70042.US  
; CURRENT APPLICATION NUMBER: US/60/479,073  
; PRIOR FILING DATE: 2003-06-17  
; NUMBER OF SEQ ID NOS: 526  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 495  
; LENGTH: 4574  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-479-073-495

Query Match 6.2%; Score 113.5; DB 7; Length 4574;  
Best Local Similarity 20.1%; Pred. No. 0.13;  
Matches 73; Conservative 65; Mismatches 169; Indels 57; Gaps 10;

QY 14 AEOFKOKAL-----RPAVDPREVLY-----YQEFDEMEASHLPYIEQOMOL 55  
DB 2013 AEEVQKSLAAEEBAARQKALAEVERLKANVEEARLREARAEQESARQLQAEAAOK 2072  
QY 56 AEOFKOKALAFASRGNHVSPTHYVESDAPPAIIPDREVLYREFDEME-EAFASR--GNH 112  
DB 2073 RLQAEKKAHAFVQK-----EQLQOOLQOESVL-----DQLGEEAARRAAE 2119  
QY 113 VSPAHYVESDASQAAPYIEQNOVIAHQEKEKVLAF--SRGNHDSPTHYVESDAKPAI 171  
DB 2120 AEEARVQAEAREAAQRQVEEAEERLKQAEQQAQAQAQAALREAEQEAARRQA 2179  
QY 172 PDREVLYREFDEMECSQHLPIEQQMMLAEQFKOKALAFASRGNHVAPTHYVESDAKPA 231  
DB 2180 EQAALROKQADAEEMEKHKFAEQTLROKQAVEDELTLRLQ-----LEETDHQKN 2230  
QY 232 LVPDKEVLYQYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEXVLA 291  
DB 2231 LL-----DEELQRLKAEATBAARQSOVEEELFSVRYQWELSKLARIEAENRA 2280  
QY 232 FASRGNHVSPTHYVESEPOVVVTPDKELIYEAPDE---MEBASKALIEGQMAEMLK 348  
DB 2281 LILRKDNTQRFLOEAEKMKQVAEEAARLSVAQEAARLRLQLAEDLAQQRALAEKMLK 2340

QY 349 SKIO 352  
DB 2341 EXMO 2344

RESULT 5  
US-60-479-073-497

Sequence 497, Application US/60479073  
GENERAL INFORMATION:  
APPLICANT: De Wilde, Gert Jules Hector  
APPLICANT: Saunders, Michael John Scott  
APPLICANT: Logghe, Marc Georges  
TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the  
TITLE OF INVENTION: prevention and/or treatment of metabolic diseases and nucleotide  
FILE REFERENCE: D00590.70042.US  
CURRENT APPLICATION NUMBER: US/60/479, 073  
CURRENT FILING DATE: 2003-06-17  
NUMBER OF SEQ ID NOS: 526  
SOFTWARE: Patentin version 3.2  
SEQ ID NO 497  
LENGTH: 4684  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-60-479-073-497

## Query Match

6.2%; Score 113.5; DB 7; Length 4684;  
Best Local Similarity 19.7%; Pred. No. 0.13;  
Matches 71; Conservative 63; Mismatches 176; Indels 51; Gaps 8;

QY 14 AEOFKOKAL-----RPVLPDREV-----YQFDEMEASHLPYIEQGMOL 55  
DB 2123 AEEVVKSLAEEBAARORKALEVERLKAKVEARSLRERAEQSARQLQAQEAOK 2182  
QY 56 AEFKOKALAFASRGNHVSTHYVPSDARPAIIPREVLRHREFDEMEAFASRGNHVSP 115  
DB 2183 RLQAEKHAFAVQOK-----EQLQOTLQOQSVDLRGRGEAA-RRAAEEAE 2232  
QY 116 AHVPESDASQAPYIEQAQVIAHQFEKVLAFASRGNHVSTHYVPSDARPAIIPDR 174  
DB 2233 ARVQAREAAQSRQVEERLQKSAEQAQARAQAAAEKLRKAEQAARAAQEAQ 2292  
QY 175 EVLYREFDEMESSOHLPYIEQGMMLAEQFKOKALAFASRGNHVAPTHYTESDAKALVP 234  
DB 2293 ALRQKQADAEEMKHKFAEQTLRQAEVQELTLRLQ-----LETDHQKLL- 2342  
QY 235 DKEVLVQOYDEMEAFASRGNHVAPTHYTESDASASLPYMETRAIAGQFEKVLAFAS 294  
DB 2343 -----DEELQRLKATEARARORSQVEEELFSVRVQMEELSKLKARIEAENRALIL 2393  
QY 295 RGNHVSPTHYVPESEPOVVTPDKELLYEAFDE---MEASKALLIEGQRMAMEMKSKI 351  
DB 2394 RKDNDTORFLQEAEMKQVAAEBAARLSVAQEAARLRQLAEEDLQAQALAEKMKEXM 2453  
QY 352 Q 352  
DB 2454 Q 2454

RESULT 6  
US-60-479-073-506

Sequence 506, Application US/60479073  
GENERAL INFORMATION:  
APPLICANT: De Wilde, Gert Jules Hector  
APPLICANT: Saunders, Michael John Scott  
APPLICANT: Logghe, Marc Georges  
INVENTION: Amino acid sequences useful for developing compounds for the  
prevention and/or treatment of metabolic diseases and nucleotide  
sequences encoding such amino acid sequences.  
US-60-479-073-506

NUMBER OF SEQ ID NOS: 526  
SOFTWARE: Patentin version 3.2  
SEQ ID NO 506  
LENGTH: 4684  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-60-479-073-506

Query Match  
6.2%; Score 113.5; DB 7; Length 4684;  
Best Local Similarity 19.7%; Pred. No. 0.13;  
Matches 71; Conservative 63; Mismatches 176; Indels 51; Gaps 8;

QY 14 AEOFKOKAL-----RPVLPDREV-----YQFDEMEASHLPYIEQGMOL 55  
DB 2123 AEEVVKSLAEEBAARORKALEVERLKAKVEARSLRERAEQSARQLQAQEAOK 2182  
QY 56 AEFKOKALAFASRGNHVSTHYVPSDARPAIIPREVLRHREFDEMEAFASRGNHVSP 115  
DB 2183 RLQAEKHAFAVQOK-----EQLQOTLQOQSVDLRGRGEAA-RRAAEEAE 2232  
QY 116 AHVPESDASQAPYIEQAQVIAHQFEKVLAFASRGNHVSTHYVPSDARPAIIPDR 174  
DB 2233 ARVQAREAAQSRQVEERLQKSAEQAQARAQAAAEKLRKAEQAARAAQEAQ 2292  
QY 175 EVLYREFDEMESSOHLPYIEQGMMLAEQFKOKALAFASRGNHVAPTHYTESDAKALVP 234  
DB 2293 ALRQKQADAEEMKHKFAEQTLRQAEVQELTLRLQ-----LETDHQKLL- 2342  
QY 235 DKEVLVQOYDEMEAFASRGNHVAPTHYTESDASASLPYMETRAIAGQFEKVLAFAS 294  
DB 2343 -----DEELQRLKATEARARORSQVEEELFSVRVQMEELSKLKARIEAENRALIL 2393  
QY 295 RGNHVSPTHYVPESEPOVVTPDKELLYEAFDE---MEASKALLIEGQRMAMEMKSKI 351  
DB 2394 RKDNDTORFLQEAEMKQVAAEBAARLSVAQEAARLRQLAEEDLQAQALAEKMKEXM 2453  
QY 352 Q 352  
DB 2454 Q 2454

RESULT 7  
US-10-273-573-10712

Sequence 10712, Application US/10273573  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc  
TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES  
FILE REFERENCE: 21272-066  
CURRENT APPLICATION NUMBER: US/10/273, 573  
CURRENT FILING DATE: 2002-10-18  
PRIOR APPLICATION NUMBER: 09/522, 929  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: 09/770, 160  
PRIOR FILING DATE: 2001-01-26  
NUMBER OF SEQ ID NOS: 10994  
SOFTWARE: Custom  
SEQ ID NO 10712  
LENGTH: 365  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-273-573-10712

Query Match  
5.8%; Score 106; DB 6; Length 365;  
Best Local Similarity 20.4%; Pred. No. 0.016;

Matches 64; Conservative 41; Mismatches 100; Indels 108; Gaps 13;

QY 69 RGNHVSPTHYVPSDAR-----PAIIPDREVLRHREFDEMEAFASRGNHVSPAHYVPS 122  
DB 2 REXSDPNIVLESIRIKHSEPSIVSESV-----RKHSPESSVLES 47  
QY 123 DASQAPYIEQAQVIAHQFE---KVLAFASRGNHVSTHYVPSDARPAIIPREVLYR 179  
DB 48 -----YVSHGSELHGRVLESYVRDKHSPESSVLES-----YVRDK----- 84

Mon Jul 28 11:16:34 2003

OY 180 EFDEMEESQHLPIYIEQGMMLAEQFKOKALAFASRGNHVAPTHVYES---DAKPAIVPDK 236  
DB 85 -----HSEPPSSVLEBSF-----VRDKHSEPPSSVLESECCSSPPAM--EK 122  
OY 237 EVLYOQYDEMEEAFA-----ASRGNHVAPTHVYE-----SDASASLPYMDT 278  
DB 123 SWAENDFPELREGFRNSVYSELQEIOTKKEVKMFKENGLDECSIRINTKCKELMEL 182  
OY 279 RAIAQOFKEXVLAFAASRGNHVSPRHYVESEPOVVVTPDKETLYEAFDEMEESKAL-I 337  
DB 183 KAKREJREBCRRFRSRRCDQLEERDIIQENSFNI-----ARQANIQI 224  
OY 338 EEOGMAMMLKSK 350  
DB 225 OEIORMFORYSLK 237

RESULT 8  
US-60-479-073-498  
; Sequence 498, Application US/60479073  
; GENERAL INFORMATION:  
; APPLICANT: De Wilde, Gert Jules Hector  
; APPLICANT: Saunders, Michael John Scott  
; APPLICANT: Logghe, Marc Georges  
; TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the  
; TITLE OF INVENTION: prevention and/or treatment of metabolic diseases and nucleotide  
; TITLE OF INVENTION: sequences encoding such amino acid sequences.  
; FILE REFERENCE: D00590.70042.US  
; CURRENT APPLICATION NUMBER: US/60/479.073  
; CURRENT FILING DATE: 2003-06-17  
; NUMBER OF SEQ ID NOS: 526  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 498  
; LENGTH: 5373  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-479-073-498

Query Match 5.7%; Score 103; DB 7; Length 5373;  
Best Local Similarity 18.5%; Pred. No. 1.4;  
Matches 78; Conservative 73; Mismatches 132; Indels 138; Gaps 19;

OY 4 IPYLEQGMMLAEQFKOKALRPAYIPREVLVYQEFDEMEESH-----LPY----- 48  
DB 3584 IPOFOQ-----ROKELKKEVMEHRLVL-----DTVNEVSRLLELVPRAREGLDKLV 3631  
OY 49 -----IEQGMMLAEQFKOKA---LAFASRGNH-----VSPTHVPE 81  
DB 3632 SDANEQYKLVSDTIGRVDEIDALIQSQYEGADALAWVATRKLMALGPRL--E 3689  
OY 82 SDARPAIIPREV---LHREFDEMEEAFAASRGN-----HVSAPHVPESDA 124  
DB 3690 ODQTTAOLQVOKAFSIDIRHKDSMDLFSHSEIFGTGCEBQKTVLOEKTESLIQYEA 3749  
OY 125 -----SOAAPYIEQAQVIAHQFKEKVLAFASRGNHSDPTHVPSDKPAIIP---DRE 175  
DB 3750 ISLNSERVARLERAOVLVNOFWE-----TYEELSPWIEETRALIAQLPSPAIDHE 3800  
OY 176 VLYPEDEMEE-----SOHLPIYIEQGMMLAEQFKOKALAFASRGNHVAPTHVYESDAKP 230  
DB 3801 QLRQOQDEMRQLRESIAEHKPHIDKLKIGPOLKE----- 3835  
OY 231 ALVPDK-EVLYOQYDEMEEAFAASRGNHVAPTHVYESDASASLPYMDETRAIAGOFKEXV 289  
DB 3836 -LNPDEGEVVEEKYOKAENMYAQIKEVQRALALDEAVSQSTQIT-----EFHDKI 3886  
OY 290 LAFASRGNHVSPRHYVESEPOVVVTPDKETLYEAFDEMEESKALIEEGRMAMMLKS 349  
DB 3887 EPMLETLLENLSRLRMP---PLIPAEVDK--IRECISDNKSAT--VELEKLOPSFEALKR 3939  
OY 350 K 350

DB 3940 R 3940

RESULT 9  
US-60-479-073-500  
; Sequence 500, Application US/60479073  
; GENERAL INFORMATION:  
; APPLICANT: De Wilde, Gert Jules Hector  
; APPLICANT: Saunders, Michael John Scott  
; APPLICANT: Logghe, Marc Georges  
; TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the  
; TITLE OF INVENTION: prevention and/or treatment of metabolic diseases and nucle  
; TITLE OF INVENTION: sequences encoding such amino acid sequences.  
; FILE REFERENCE: D00590.70042.US  
; CURRENT APPLICATION NUMBER: US/60/479.073  
; CURRENT FILING DATE: 2003-06-17  
; NUMBER OF SEQ ID NOS: 526  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 500  
; LENGTH: 5430  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-479-073-500

Query Match 5.7%; Score 103; DB 7; Length 5430;  
Best Local Similarity 18.5%; Pred. No. 1.4;  
Matches 78; Conservative 73; Mismatches 132; Indels 138; Gaps 19;

OY 4 IPYLEQGMMLAEQFKOKALRPAYIPREVLVYQEFDEMEESH-----LPY----- 48  
DB 3640 IPOFOQ-----ROKELKKEVMEHRLVL-----DTVNEVSRLLELVPRAREGLDKLV 3687  
OY 49 -----IEQGMMLAEQFKOKA---LAFASRGNH-----VSPTHVPE 81  
DB 3688 SDANEQYKLVSDTIGRVDEIDALIQSQYEGADALAWVATRKLMALGPRL--E 3745  
OY 82 SDARPAIIPREV---LHREFDEMEEAFAASRGN-----HVSAPHVPESDA 124  
DB 3746 ODQTTAOLQVOKAFSIDIRHKDSMDLFSHSEIFGTGCEBQKTVLOEKTESLIQYEA 3805  
OY 125 -----SOAAPYIEQAQVIAHQFKEKVLAFASRGNHSDPTHVPSDKPAIIP---DRE 175  
DB 3806 ISLNSERVARLERAOVLVNOFWE-----TYEELSPWIEETRALIAQLPSPAIDHE 3856  
OY 176 VLYPEDEMEE-----SOHLPIYIEQGMMLAEQFKOKALAFASRGNHVAPTHVYESDAKP 230  
DB 3857 QLRQOQDEMRQLRESIAEHKPHIDKLKIGPOLKE----- 3891  
OY 231 ALVPDK-EVLYOQYDEMEEAFAASRGNHVAPTHVYESDASASLPYMDETRAIAGOFKEXV 289  
DB 3892 -LNPDEGEVVEEKYOKAENMYAQIKEVQRALALDEAVSQSTQIT-----EFHDKI 3942  
OY 290 LAFASRGNHVSPRHYVESEPOVVVTPDKETLYEAFDEMEESKALIEEGRMAMMLKS 349  
DB 3943 EPMLETLLENLSRLRMP---PLIPAEVDK--IRECISDNKSAT--VELEKLOPSFEALKR 3995  
OY 350 K 350  
DB 3996 R 3996

RESULT 10  
US-10-333-120A-10  
; Sequence 10, Application US/10333120A  
; GENERAL INFORMATION:  
; APPLICANT: Januliczky, Robert  
; APPLICANT: Taneli, Francesco  
; APPLICANT: Sjolholm, Anders  
; APPLICANT: Pozzi, Gianni  
; APPLICANT: Bjorck, Lars  
; TITLE OF INVENTION: PROTEIN  
; FILE REFERENCE: 13519PCTUS  
; CURRENT APPLICATION NUMBER: US/10/333,120A



CURRENT FILING DATE: 2002-01-14  
PRIOR APPLICATION NUMBER: PCT/EP01/08409  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: SE 0002728-4  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 693  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-10-333-120A-10

Query Match 5.2%; Score 95; DB 6; Length 693;  
Best Local Similarity 22.0%; Pred. No. 0.38;  
Matches 55; Conservative 49; Mismatches 122; Indels 24; Gaps 7;

QY 15 EGFKQKALRAVIFPDRVLYOEFDME--EASHLYIQGQMLAEQFKALAFASRQ 70  
DB 230 ELVKVKNRPR--DEQKIQAEAEVESKQAEATRLKKTIDREAE--EAKRRADAKEQ 285  
QY 71 NHVSPHYVPESDAPAIIPDREVLHREFDEMEAFASRGNHVSFAHYV--PESDASQAP 129  
DB 286 GK-----PKGRAKRGVPELATPDCKENDAKSSDSSVGEETLPSPLKPEKKVLAEEK 338  
QY 130 YIQAOVIYAHQFEKVLAFASRGNHDSPTHYVPSDAKPAIIPDREVLHREFDEMEESQH 189  
DB 339 KVEAKKAKEDQEE-----DRRNYPTNYKTLELEIAE--DVEYKAELELVKEEAK 390  
QY 190 LPYIEQGMMLAEQFKQKALAFASRGNHVAPTHYVTESDAKPALVPDEVLVYQYDMEEA 249  
DB 391 EPRNEEKVQAKAEVESKQAEATRLKKTIDRKAKEEAKKAAEDVKKEKPAEQOPA 450  
QY 250 FASRGNHVP 259  
DB 451 PAPKAEKPA 460

RESULT 11  
US-10-391-363A-61  
Sequence 61, Application US/10391363A  
GENERAL INFORMATION:  
APPLICANT: Monsanto Technology LLC  
APPLICANT: Valentin, Henry E  
APPLICANT: Venkatesh, Tyamagondlu V  
APPLICANT: Karunananda, Balasulojini  
TITLE OF INVENTION: Homogentisate Prenyl Transferase ("HPT") Nucleic Acids and Polypeptide  
TITLE OF INVENTION: and Uses Thereof  
FILE REFERENCE: REN-02-052 MON-52933  
CURRENT FILING DATE: 2003-03-18  
PRIOR APPLICATION NUMBER: US/10391,363A  
PRIOR FILING DATE: 2002-03-19  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 61  
LENGTH: 970  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-391-363A-61

Query Match 5.2%; Score 95; DB 6; Length 970;  
Best Local Similarity 21.4%; Pred. No. 0.62;  
Matches 72; Conservative 47; Mismatches 123; Indels 94; Gaps 14;

31 EVLYOEFDMEASHLYIQGQ-----QLAEQFKQKALAFASRGNHVSPTHYV 80  
VQOQHLDNV--SSHLYIIRSYKEASLDLKPRAEOIOLSTRLEARNASTNLVYK 323  
PAIIPDREVLHREFDE--MEAFASRGNHVS-----PAHYVE 121  
KPEKTEKALEEAEIKPEPSTNLKAVTLDEQTAREAPE 383

QY 122 S-----DAS--QAAPYIEQVIAHQFE-----KYLAFASRGNHDSPTHYV 162  
DB 384 DKLVSKVDAAMVKAAYKASEAIDNVNEBGLAKQKLEIFGEFIIKEGDNNSADDLSE 443  
QY 163 ESDAKPAIIPDREVLHREFDEMEESQHPY-----IEGMMMLAEQFKQKALAFASRGN 215  
DB 444 RTTQKAPVYVMDAEAEEDSVTKDRTTYETTRGLEIENEMVIDETKST----- 495  
QY 216 HVAPTHYVTESDAKPALVPDKEVLVYQYDMEEAFAFRGNHVAPTHYVTESD----- 267  
DB 496 -----ETKTEAPSRIVMDKE-----GDEETKSTETETETAPSRIVMETEDEETMNS 542  
QY 268 -ASASLPYMDETRAIQREKVLAFASRGNHVS 302  
DB 543 RARAS-----DEAEALS--KSSQVASMELISQSPR 571

RESULT 12  
PCT-US02-18638A-2  
Sequence 2, Application PC/TUS0218638A  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc. et al.  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
TITLE OF INVENTION: OF CERVICAL CANCER  
FILE REFERENCE: MRI-035PC  
CURRENT FILING DATE: 2002-06-12  
PRIOR APPLICATION NUMBER: PCT/US02/18638A  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/298,159  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/298,155  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/335,936  
PRIOR FILING DATE: 2001-11-14  
NUMBER OF SEQ ID NOS: 238  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 3907  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US02-18638A-2

Query Match 5.2%; Score 95; DB 1; Length 3907;  
Best Local Similarity 19.3%; Pred. No. 4.6;  
Matches 68; Conservative 54; Mismatches 145; Indels 86; Gaps 14;

QY 19 QKALRAVIFPDRVLYOEFDMEEA--SHLYIQGQMLAEQFKQKALAFASRGNHVSPT 76  
DB 2020 QKQVLALEI--DVEQVSRFIELEQKNTLMRLROONQALEKMRKFTDEQ--AIDRE 2077  
QY 77 HVPESDAPAIIPDREVLHREFDEMEEAFAFRGNHVSFAHYV--PESDASQAPYIEQV 136  
DB 2078 H-----ERDVFOQETQKLEQOL-----KVP--RQPISEHQ-----TREVEQ 2113  
QY 137 IAHQFEKVLAFASRGNHDSPTHYVPSDAKPAIIPDREVLHREFDEMEESQHPYIEQG 196  
DB 2114 LANHLKEX-----IDKCELLLSKEQLORDQENNE-----IEVL 2149  
QY 197 MMLAEQFKQKALAFASRGNHVAPTHYVTESDAKPALV-----PDKEV--LY 240  
DB 2150 EFRVLELEQALLVSADTEQKVEDRKHFGAVEAKPELSLEVQLAERDAIDRKEKEITNLE 2209  
QY 241 QOYDMEEAFAFRGNHVAPTHYVTESDASASLPYMDETRAIQREKVL--LAFASRGNH 298  
DB 2210 EOLEQFRELEKNEEVOOLHMOLEIOTKEKSTTRLOEBOENKLRKDMKEKGLAIK--- 2266  
QY 299 VSPRAHYVESQPVVTPDKETLYEAFDEM--EASAAALIEGQRAEMMLK 348  
DB 2267 -----ESDANSTODQHVLFGKFAOIIQEKVEIDQNEQVTKLOQOLK 2309

RESULT 13  
PCT-US02-18638A-6

```

;
; TYPE: PRI
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: DOMAIN

```

13 LAEPFKKRAIPEAVIPDRFVYOFENEFKASHIPEYFOCM-----OIAEFKRAIAF 66  
Best. Local Similarity 20.3%; Pred. NO. 2.1;  
Matches 74; Conservative 61; Mismatches 141; Indels 85; Gaps 17;

13 LAAEFOKAKALRAVPIPREVLVOEFDEMEESHLPRIEOM-----QLAEOFKAKALAF- 66  
530 LKQOAKKRAQO-----LAOTLOOEOGS-----OGIRHOVEQLSSLSLKQEOOLK 574

```

QY 67 -----ASRONHVSPTHYVESDARPAIIPDREVLHREPDMEEAFAFRGNHVS PAHYV 119
Db 575 EVAEKQDXTRODHAQ--OLATAAEEREA SLRERDALKQLEALEKEXAKLEILQOQLQV 632
QY 120 PESDASQAAPYIEQAQVIAHQFKEKV-----LAFASRGNHDSPTHYV-----ESDAKP 168
Db 633 ANEARDGNOTSVTQHQREKAE LSRKVEELQACVETARQOEHOEAQVAELELQLRSEQOK 692
QY 169 AIPDREVLHREPDMEESQHLPYIEQGM-----MLAQFKOKALAFASRGNHVAPTHYV 223
Db 693 A--TEKERVNAQEKQLOE--QLOALKESLKYTKGSLBEKRRADALEEQRCISELKAE 748
QY 224 TESDAKPALVPDKVLYQOYDEMEEAFAFRGNHVAPTHYVESDASASLPYMDETRAIAG 283
Db 749 TRS-----LVEOHK--RERKELEERAGR-----KGLFARLQOLGEAHOAET 788
QY 284 QFKEXVLAFAFRGNHVS PRHYVPESEPOVVVTPDKEL-----YEAFDMEEASKALLIE 338
Db 789 EVLRRELAEMAAQHTA-----ESECEQLV--KEVAAMRERYED--SQOEAOYGAMFQ 838
QY 339 E 339
Db 839 E 839

```

Search completed: July 25, 2003, 13:20:07  
 Job time : 15 secs

FILING DATE: 25-Jan-2000

Mon Jul 22

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/921,887  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: WARREN, WILLIAM L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 03063-0380  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 352 amino acids  
TYPE: amino acid  
STRANDEDNESS: Not Relevant  
TOPOLOGY: Not Relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Hepatitis virus  
SEQUENCE DESCRIPTION: SEQ ID NO: 52:  
US-09-491-146-52

Query Match 100.0%; Score 1818; DB 18; Length 352;  
Best Local Similarity 100.0%; Pred. No. 3,1e-156;  
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAHIFLEOGMHLAEQFKOKALRPVAPIDREVLYOEFDMEEAASHLPYIEOGMQLAEQFK 60  
DB 1 AAHIFLEOGMHLAEQFKOKALRPVAPIDREVLYOEFDMEEAASHLPYIEOGMQLAEQFK 60  
QY 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLHREDEMEAFASRGNHVSPTAHYVP 120  
DB 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLHREDEMEAFASRGNHVSPTAHYVP 120  
QY 121 ESDASQAAPYIEQAOVIAHQFKEKVLAFASRGNHDSPTHYVESDAPKPAIIPDREVLHRE 180  
DB 121 ESDASQAAPYIEQAOVIAHQFKEKVLAFASRGNHDSPTHYVESDAPKPAIIPDREVLHRE 180  
QY 181 FDEMEESQHLPYIEOGMHLAEQFKOKALAFASRGNHVAPTHYVETSDAKPALVPDKEVLY 240  
DB 181 FDEMEESQHLPYIEOGMHLAEQFKOKALAFASRGNHVAPTHYVETSDAKPALVPDKEVLY 240  
QY 241 QOYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300  
DB 241 QOYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300  
QY 301 PRHYVESEPQVVVTPDKETILYFAFDMEEASKALIEGQMAEMLSKIQ 352  
DB 301 PRHYVESEPQVVVTPDKETILYFAFDMEEASKALIEGQMAEMLSKIQ 352

## RESULT 2

US-09-491-146a-52  
Sequence 52, Application US/09491146A  
GENERAL INFORMATION:  
APPLICANT: The Government of the United States of America, as represented by  
APPLICANT: The Secretary, Department of Health and Human Services,  
APPLICANT: C/o Centers for Disease Control and Prevention  
APPLICANT: Khudyakov, Yuri  
APPLICANT: Fields, Howard  
TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION ENDONUCLEASE ASSISTED  
FILE REFERENCE: 14114.0344U2  
CURRENT APPLICATION NUMBER: US/09/491,146A  
PRIOR FILING DATE: 2000-01-25  
PRIOR APPLICATION NUMBER: 08/921,887  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 52

LENGTH: 352  
TYPE: PRP  
ORGANISM: Hepatitis C virus  
US-09-491-146a-52

Query Match 100.0%; Score 1818; DB 18; Length 352;  
Best Local Similarity 100.0%; Pred. No. 3,1e-156;  
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAHIFLEOGMHLAEQFKOKALRPVAPIDREVLYOEFDMEEAASHLPYIEOGMQLAEQFK 60  
DB 1 AAHIFLEOGMHLAEQFKOKALRPVAPIDREVLYOEFDMEEAASHLPYIEOGMQLAEQFK 60  
QY 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLHREDEMEAFASRGNHVSPTAHYVP 120  
DB 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLHREDEMEAFASRGNHVSPTAHYVP 120  
QY 121 ESDASQAAPYIEQAOVIAHQFKEKVLAFASRGNHDSPTHYVESDAPKPAIIPDREVLHRE 180  
DB 121 ESDASQAAPYIEQAOVIAHQFKEKVLAFASRGNHDSPTHYVESDAPKPAIIPDREVLHRE 180  
QY 181 FDEMEESQHLPYIEOGMHLAEQFKOKALAFASRGNHVAPTHYVETSDAKPALVPDKEVLY 240  
DB 181 FDEMEESQHLPYIEOGMHLAEQFKOKALAFASRGNHVAPTHYVETSDAKPALVPDKEVLY 240  
QY 241 QOYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300  
DB 241 QOYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300  
QY 301 PRHYVESEPQVVVTPDKETILYFAFDMEEASKALIEGQMAEMLSKIQ 352  
DB 301 PRHYVESEPQVVVTPDKETILYFAFDMEEASKALIEGQMAEMLSKIQ 352

## RESULT 3

US-09-881-239-5  
Sequence 5, Application US/09881239  
GENERAL INFORMATION:  
APPLICANT: CHEN, David Y.  
APPLICANT: ARCANDEL, Phillip  
APPLICANT: TANDESKE, Laura  
APPLICANT: GEORGE-NASCIMENTO, Carlos  
APPLICANT: COIT, Doris  
APPLICANT: MEDINA-SELBY, Angelica  
TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY  
FILE REFERENCE: 2302-16073 / PP16073.003  
CURRENT APPLICATION NUMBER: US/09/881,239  
PRIOR FILING DATE: 2001-06-14  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 829  
TYPE: PRP  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: MEPA 12  
US-09-881-239-5

Query Match 28.2%; Score 512.5; DB 23; Length 829;  
Best Local Similarity 62.8%; Pred. No. 6,8e-37;  
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

QY 167 KPAILPDBEVLYREFDEMEE-SQHLPYIEOGMHLAEQFKOKALAFASRGNHVAPTHYVTE 225  
DB 167 KPAILPDBEVLYREFDEMEECSQHLPYIEOGMHLAEQFKOKALG-SRG----- 444  
QY 226 SPARKALVPDKEVLYQOYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGQF 285  
DB 226 SPARKALVPDKEVLYQOYDEME-----CSQAAPYIEQAOVIAHQF 483  
QY 286 KEKVLAFASRGNHVSPTHYVESSEPQVVVTPDKETILYFAFDMEE-ASKAALIEGQMA 344  
DB 286 KEKVLAFASRGNHVSPTHYVESSEPQVVVTPDKETILYFAFDMEEASKAALIEGQMA 344  
QY 484 KEKVLGLI-----DNDQVVVTPDKETILYFAFDMEEASKAALIEGQMA 529

QY 345 EMLKSKIQ 352  
Db 530 EMLKSKIQ 537

RESULT 4  
US-09-791-537-91118  
; Sequence 91118, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 91118  
; LENGTH: 839  
; TYPE: PRT  
; ORGANISM: synthetic construct  
US-09-791-537-91118

Query Match 28.2%; Score 512.5; DB 22; Length 839;  
Best Local Similarity 62.8%; Pred. No. 6.9e-37;  
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

QY 167 KPAIIPDREVLVREFPEMEE-SOHLPIYIEGMMLEAFQFKALAFASRGNHVAPTHYVTE 225  
Db 489 KPAIIPDREVLVREFPEMEECSOHLPIYIEGMMLEAFQFKALGL-SRG----- 536  
QY 226 SPAKPLVPDKKVLVYQYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGOF 285  
Db 537 --GKPAIVDPKVLVYQYDEME-----CSQAAPYIEQAQVIYAHOF 575  
QY 286 KEKVLAFASRGNHVSPRHYVPESEPOVVTTPDKELIYEAFFDEME-ASKAALIEEGORMA 344  
Db 576 KEKVLGLI-----DNDQVVVTPDKELIYEAFFDEMECASQAALIEEGORMA 621  
QY 345 EMLKSKIQ 352  
Db 622 EMLKSKIQ 629

RESULT 5  
US-09-881-654-4  
; Sequence 4, Application US/09881654  
; GENERAL INFORMATION:  
; APPLICANT: CHEN, David Y.  
; APPLICANT: ARCANDEL, Phillip  
; APPLICANT: TANDESKE, Laura  
; APPLICANT: GEORGE-NASCIENTO, Carlos  
; APPLICANT: COLT, Doris  
; APPLICANT: MEDINA-SELBY, Angelica  
; TITLE OF INVENTION: IMMUNOSSAYS FOR ANTI-HCV ANTIBODIES  
; FILE REFERENCE: 2302-17039 / PPI7039,002  
; CURRENT APPLICATION NUMBER: US/09/881,654  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: 60/212,082  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 60/280,811  
; PRIOR FILING DATE: 2001-04-02  
; APPLICATION NUMBER: 60/280,867  
; Filing DATE: 2001-04-02  
; NOS: 7  
; 2.0

FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: MEPA 7.1  
US-09-881-654-4

Query Match 28.2%; Score 512.5; DB 23; Length 1099;  
Best Local Similarity 62.8%; Pred. No. 1.1e-36;  
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

QY 167 KPAIIPDREVLVREFPEMEE-SOHLPIYIEGMMLEAFQFKALAFASRGNHVAPTHYVTE 225  
Db 701 KPAIIPDREVLVREFPEMEECSOHLPIYIEGMMLEAFQFKALGL-SRG----- 748  
QY 226 SPAKPLVPDKKVLVYQYDEMEAFASRGNHVAPTHYVESDASASLPYMDETRAIAGOF 285  
Db 749 --GKPAIVDPKVLVYQYDEME-----CSQAAPYIEQAQVIYAHOF 787  
QY 286 KEKVLAFASRGNHVSPRHYVPESEPOVVTTPDKELIYEAFFDEME-ASKAALIEEGORMA 344  
Db 788 KEKVLGLI-----DNDQVVVTPDKELIYEAFFDEMECASQAALIEEGORMA 833  
QY 345 EMLKSKIQ 352  
Db 834 EMLKSKIQ 841

RESULT 6  
US-08-417-478-1  
; Sequence 1, Application US/08417478  
; GENERAL INFORMATION:  
; APPLICANT: CASEY, JAMES M.  
; APPLICANT: BODE, SUZANNE L.  
; APPLICANT: ZECK, BILLY J.  
; APPLICANT: YAMAGUCHI, JULIE  
; APPLICANT: FRAIL, DONALD E.  
; APPLICANT: DESAI, SURESH M.  
; APPLICANT: DEVARE, SUSHIL G.  
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
; STREET: ONE ABBOTT PARK ROAD  
; CITY: ABBOTT PARK  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/417,478  
; FILING DATE:  
; CLASSIFICATION: 435  
; APPLICATION DATA:  
; APPLICATION NUMBER: US/08/144,099  
; FILING DATE:  
; APPLICATION NUMBER: US 07/830,024  
; FILING DATE: 01-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FOREMSKI, PRISCILLA E.  
; REGISTRATION NUMBER: 33,207  
; REFERENCE/DOCKET NUMBER: 5131, US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 708-937-9556  
; TELEFAX: 708-937-6365  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3011 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

Mon Jul 28 11:16:34 2003



Sequence 10, Application US/07635451

## GENERAL INFORMATION:

APPLICANT: OKAYAMA, Hiroto

APPLICANT: FUKU, Isao

APPLICANT: MORI, Chisato

APPLICANT: TAKAMIZAWA, Akahisa

APPLICANT: YOSHIDA, Iwao

TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESSEE: Armstrong, Nikaido, Marmelstein, Kubovcik, &amp;

ADDRESS: Murray

STREET: 1725 K St. N.W. Suite 1000

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/635,451

FILING DATE: 19901228

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 2-167466

FILING DATE: 25-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 2-230921

FILING DATE: 31-AUG-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 2-305605

FILING DATE: 09-NOV-1990

ATTORNEY/AGENT INFORMATION:

NAME: O'Toole, J. Herbert

REGISTRATION NUMBER: 31,404

REFERENCE/DOCKET NUMBER: 900703

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 659-2930

TELEFAX: (202) 887-0357

TELEX: 440142

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 2013 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-635-451-10

Query Match 14.1%; Score 256; DB 3; Length 2013;

Best Local Similarity 31.1%; Pred. No. 5.8e-13;

Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

DB 23 RPAIVPDRVLYOEFDMEEM--ASHLPYIEGQWLAOFKOKALAF--ASRONHVSPTHY 78

DB 1691 RPAIVPDRVLYOEFDMEEM--ASHLPYIEGQWLAOFKOKALAF--ASRONHVSPTHY 78

DB 79 VPESDARPAIIPDRVLYOEFDMEEM--ASHLPYIEGQWLAOFKOKALAF--ASRONHVSPTHY 133

DB 1746 ----AAP-----VVEKSKRALETFWAKHMMNFTSGIOYLAGLSTLPGNPAIASLMAF 1793

DB 134 -AOVIAHOFKEKYLAF-----ASRGHNDSPHYVPESDARPAI-----AE 170

DB 794 TASTITSLTOSTLFLNIGWAAOLAPSAFAVAGIAGAAVGSIGLGKVLVDIILA 1853

DB 1691 RPAIVPDRVLYOEFDMEEM--ASHLPYIEGQWLAOFKOKALAF--ASRONHVSPTHY 78

DB 79 VPESDARPAIIPDRVLYOEFDMEEM--ASHLPYIEGQWLAOFKOKALAF--ASRONHVSPTHY 133

DB 1746 ----AAP-----VVEKSKRALETFWAKHMMNFTSGIOYLAGLSTLPGNPAIASLMAF 1793

DB 1914 QMMNRLIAFASRGHNSPTHYVPESDA 1940

RESULT 10

US-07-769-996B-12

Sequence 12, Application US/07769996B

GENERAL INFORMATION:

APPLICANT: OKAYAMA, Hiroto

APPLICANT: FUKU, Isao

APPLICANT: MORI, Chisato

APPLICANT: TAKAMIZAWA, Akahisa

APPLICANT: YOSHIDA, Iwao

TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Armstrong, Nikaido, Marmelstein

ADDRESS: Murray

STREET: 1725 K St. N.W. Suite 1000

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/769,996B

FILING DATE: 19911002

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 2-167466

FILING DATE: 25-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 2-230921

FILING DATE: 31-AUG-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 2-305605

FILING DATE: 09-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/635,451

FILING DATE: 28-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: O'Toole, J. Herbert

REGISTRATION NUMBER: 31,404

REFERENCE/DOCKET NUMBER: 900703B

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 659-2930

TELEFAX: (202) 887-0357

TELEX: 440142

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 2013 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-769-996B-12

Query Match 14.1%; Score 256; DB 3; Length 2013;

Best Local Similarity 31.1%; Pred. No. 5.8e-13;

Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

DB 23 RPAIVPDRVLYOEFDMEEM--ASHLPYIEGQWLAOFKOKALAF--ASRONHVSPTHY 78

DB 1691 RPAIVPDRVLYOEFDMEEM--ASHLPYIEGQWLAOFKOKALAF--ASRONHVSPTHY 78

DB 79 VPESDARPAIIPDRVLYOEFDMEEM--ASHLPYIEGQWLAOFKOKALAF--ASRONHVSPTHY 133

DB 1746 ----AAP-----VVEKSKRALETFWAKHMMNFTSGIOYLAGLSTLPGNPAIASLMAF 1793



OY 134 -AQTIAHQEKELVAF-----ASRGNDSPTHVYVESDAKPAI-----170  
DB 1794 TASITSPLTQSTLFLNIGGVAAQLAPPSAASAFVAGAGAAVSGIGKVLVDILA 1953  
OY 171 -----IPDREVLVREFDEMEESQHL--PYIEQGMVL-----AE 201  
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGECAV 1913  
OY 202 QFKOKALAFASRGNHVAPTHVYTESDA 228  
DB 1914 QMNRRLIAFASRGNHVSPTHVYVESDA 1940

RESULT 11  
US-07-769-996C-12  
Sequence 12, Application US/07769996C  
GENERAL INFORMATION:  
APPLICANT: OKAYAMA, Hiroto  
APPLICANT: FUKU, Isao  
APPLICANT: MORI, Chisato  
APPLICANT: TAKAMIZAWA, Akahisa  
APPLICANT: YOSHIDA, Iwao  
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC  
TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Armstrong, Westerman, Hattori, Mclelland  
STREET: 1725 K St. N.W. Suite 1000  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/769,996C  
FILING DATE: 02-OCT-1991  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-167466  
FILING DATE: 25-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-230921  
FILING DATE: 31-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-305605  
FILING DATE: 09-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/635,451  
FILING DATE: 28-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Mclelland, Le-Nhung  
REGISTRATION NUMBER: 31,541  
REFERENCE/DOCKET NUMBER: 900703B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 659-2930  
TELEFAX: (202) 887-0357  
TELEX: 440142  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2013 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-769-996C-12

Query Match 14.1%; Score 256; DB 3; Length 2013;  
Best Local Similarity 31.1%; Pred. No. 5.8e-13;  
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

OY 23 RPAVDPREVLVYOEFEDEME-ASHLPYIEQGMVLAQFQKOKALAF-----ASRGNDSPTHV 78  
DB 1691 RPAVDPREVLVYOEFEDEMECAHLPYIEQGMVLAQFQKOKALGLLQTNKQEA-----1745  
OY 79 VPESDARPAIIPDREVLVREFDEMEEAFASR-GNHVSPAHVYVESDASQAAPYIEQ-----133  
DB 1746 -----AAP-----VSESKRALETFMAKHMNFIISGIQYLAGLSTLPCNPAAIATLMVF 1793  
OY 134 -AQTIAHQEKELVAF-----ASRGNDSPTHVYVESDAKPAI-----170  
DB 1794 TASITSPLTQSTLFLNIGGVAAQLAPPSAASAFVAGAGAAVSGIGKVLVDILA 1853  
OY 171 -----IPDREVLVREFDEMEESQHL--PYIEQGMVL-----AE 201  
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGECAV 1913  
OY 202 QFKOKALAFASRGNHVAPTHVYTESDA 228  
DB 1914 QMNRRLIAFASRGNHVSPTHVYVESDA 1940

RESULT 12  
US-08-099-706-12  
Sequence 12, Application US/08099706  
GENERAL INFORMATION:  
APPLICANT: OKAYAMA, Hiroto  
APPLICANT: FUKU, Isao  
APPLICANT: MORI, Chisato  
APPLICANT: TAKAMIZAWA, Akahisa  
APPLICANT: YOSHIDA, Iwao  
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC  
TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Armstrong, Nikaido, Marmelestein  
ADDRESSEE: Kubovcik & Murray  
STREET: 1725 K St. N.W. Suite 1000  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/099,706  
FILING DATE: 30-JUL-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-167466  
FILING DATE: 25-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-230921  
FILING DATE: 31-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-305605  
FILING DATE: 09-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/635,451  
FILING DATE: 28-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/769,996B  
FILING DATE: 02-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Toole, J. Herbert  
REGISTRATION NUMBER: 31,404  
REFERENCE/DOCKET NUMBER: 900703B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 659-2930  
TELEFAX: (202) 887-0357

TELEX: 440142  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2013 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-099-706-12

Query Match 14.1%; Score 256; DB 4; Length 2013;  
Best Local Similarity 31.1%; Pred. No. 5.8e-13;  
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPAVDPREVLYGFDEMEEASHLPYIEGQMLAEQKOKALAF--ASRQNHVSPTHY 78  
DB 1691 RPAIVDPRELLYGFDEMEECASHLPYIEGQMLAEQKOKALGLQTATKQAE----- 1745  
QY 79 VPESDARPAIIPREVLYHREFDEMEERAFSR-GNHVSPAHYVPESDASQAPYIEQ---- 133  
DB 1746 ----AAP-----VVESEKRALETFWAKHMMNFISGLOYLAGLSTLPGNPALIASLMAF 1793  
QY 134 -AQVIAHQFEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170  
DB 1794 TASITSPLTQSTLLFNILGQVAAQLAPPSASAFVAGAGIAGAAVSGIGKVLVDILA 1853  
QY 171 -----IPREVLYHREFDEMEESQHL--PYIEGQML-----AE 201  
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISGALVGVCAALIRRHVGPEGAV 1913  
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228  
DB 1914 QMMNRLIAFASRGNHVSPTHYVTESDA 1940

RESULT 13  
US-08-383-682-10  
Sequence 10, Application US/08383682  
GENERAL INFORMATION:  
APPLICANT: OKAYAMA, Hiroko  
APPLICANT: FUKU, Isao  
APPLICANT: MORI, Chisato  
APPLICANT: TAKAMIZAWA, Akahisa  
APPLICANT: YOSHIDA, Iwao  
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS PARTICLES  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Armstrong, Nikolaide, Marmelstein, Kubovcik, &  
STREET: 1725 K St. N.W. Suite 1000  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/383,682  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/720,961  
FILING DATE: 25-JUN-1991  
APPLICATION NUMBER: JP 2-167466  
DATE: 25-JUN-1990  
ON DATA: JP 2-230921

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/635,451  
FILING DATE: 28-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Toole, J. Herbert  
REGISTRATION NUMBER: 31,404  
REFERENCE/DOCKET NUMBER: 900703A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 659-2930  
TELEFAX: (202) 887-0357  
TELEX: 440142  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2013 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-383-682-10

Query Match 14.1%; Score 256; DB 7; Length 2013;  
Best Local Similarity 31.1%; Pred. No. 5.8e-13;  
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPAVDPREVLYGFDEMEEASHLPYIEGQMLAEQKOKALAF--ASRQNHVSPTHY 78  
DB 1691 RPAIVDPRELLYGFDEMEECASHLPYIEGQMLAEQKOKALGLQTATKQAE----- 1745  
QY 79 VPESDARPAIIPREVLYHREFDEMEERAFSR-GNHVSPAHYVPESDASQAPYIEQ---- 133  
DB 1746 ----AAP-----VVESEKRALETFWAKHMMNFISGLOYLAGLSTLPGNPALIASLMAF 1793  
QY 134 -AQVIAHQFEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170  
DB 1794 TASITSPLTQSTLLFNILGQVAAQLAPPSASAFVAGAGIAGAAVSGIGKVLVDILA 1853  
QY 171 -----IPREVLYHREFDEMEESQHL--PYIEGQML-----AE 201  
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISGALVGVCAALIRRHVGPEGAV 1913  
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228  
DB 1914 QMMNRLIAFASRGNHVSPTHYVTESDA 1940

RESULT 14  
US-10-085-476-2  
Sequence 2, Application US/10085476  
GENERAL INFORMATION:  
APPLICANT: De Francesco, Raffaele  
APPLICANT: Tomei, Lucia  
APPLICANT: Behrens, Sven-Erik  
TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE  
TITLE OF INVENTION: RNA-DEPENDENT RNA POLYMERASE AND TERMINAL NUCLEOTIDYL  
TRANSFERASE ACTIVITIES ENCODED BY HEPATITIS C VIRUS (HCV)  
FILE REFERENCE: IT0002PCA  
CURRENT APPLICATION NUMBER: US/10/085,476  
PRIOR FILING DATE: 2002-02-27  
PRIOR APPLICATION NUMBER: 08/952,981  
PRIOR FILING DATE: 1998-03-23  
PRIOR APPLICATION NUMBER: PCT/IT96/00106  
PRIOR FILING DATE: 1996-05-24  
PRIOR APPLICATION NUMBER: RM95A000343  
PRIOR FILING DATE: 1995-05-25  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 2201  
TYPE: PRT  
ORGANISM: CDNA clone pCD (38-9.4)  
US-10-085-476-2

Query Match 14.1%; Score 256; DB 26; Length 2201;  
Best Local Similarity 31.1%; Pred. No. 6.7e-13;

Mon Jul 28 11:16:34

RESULT 15  
 US-07-769-996B-32  
 Sequence 32, Application US/07769996B  
 GENERAL INFORMATION:  
 APPLICANT: OKAYAMA, Hiroo  
 APPLICANT: FUKE, Isao  
 APPLICANT: MORI, Chisato  
 APPLICANT: TAKAMIZAWA, Akahisa  
 APPLICANT: YOSHIDA, Iwao  
 TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS  
 TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE  
 NUMBER OF STRUCTURES: 10

COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/769,996B  
FILING DATE: 19911002  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-167466  
FILING DATE: 25-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-230921  
FILING DATE: 31-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-305605  
FILING DATE: 09-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/635,451  
FILING DATE: 28-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Toole, J. Herbert  
REGISTRATION NUMBER: 31,404  
REFERENCE/DOCKET NUMBER: 900703B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 659-2930  
TELEFAX: (202) 887-0357  
TELEX: 440142  
INFORMATION FOR SEQ ID NO: 32:

Query Match	14.1%	Score 256;	DB 3;	Length 2620;
Best Local Similarity	31.1%;	Pred. No. 8.8e-13;		
Matches 83;	Conservative 31;	Mismatches 75;	Indels 78;	Gaps 10;

OY	23	RPAVIREVLXOEFDEMEE--ASLPIETEOGMOLEQKOKALAF---ASRONHVSPTHY	78
Db	1302	RPAIVPRELLXOEFDEMEECASHLPITIEGMOLAEFPKOKALLLOTATKOAEA----	1366
OY	79	VPESDARPAIIIPREVLRHEFDEMEEAFASR-GNHVSAHYVPESDASQAAPYIO----	133
Db	1357	--AAP-----VESKRALETFTWAGHMNFISGIQYLGLSTLPNGNPATIASIMAF	1404
OY	134	-AOVIAHQFEKVLA-----ASGRNHDSPHYVPESDAKPAL-	170
Db	1405	TASITSEPTTQSLLFNILGGWAALOAPPSAAAFAGACIAGAAGVISIGLGKLVLDILA	1464
OY	171	-----IPDEVLYREFDEMEESCHL--PYIEQGML-:::--	AE 201
Db	1465	GYGAVGAALVAEKMSGEMPTREDLVNLLPALISPCALVVGVCAILRRHVGPGECAV	1524
OY	202	QFKOKALAFASRGNHVAPHYVTESDA	228
Db	1525	QMNRRLIAFSARGNHVSPHYVPESDA	1551

Search completed: July 25, 2003, 13:19:47  
Job time : 378 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 13:06:21 / Search time 97 Seconds

(without alignments)  
936.438 Million cell updates/sec

Title: US-09-491-146a-52

Perfect score: 1818  
Sequence: 1 AAH1PYLEQGHLEDFKOK.....KAALIEGQRMENLKSIIQ 352

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

S-sched: 830525 seqs, 258052604 residues

1 number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database:

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	268	14.7	3008	12	039929 hepatitis c
2	268	14.7	3010	12	093363 hepatitis c
3	268	14.7	3010	12	081825 hepatitis c
4	268	14.7	3010	12	068285 hepatitis c
5	263	14.5	3010	12	0901X1 hepatitis c
6	263	14.5	3010	12	068949 hepatitis c
7	261	14.4	3010	12	09DTE2 hepatitis c
8	261	14.4	3010	12	0933H9 hepatitis c
9	260	14.3	3010	12	0901X2 hepatitis c
10	260	14.3	3010	12	093016 hepatitis c
11	260	14.3	3010	12	0933G7 hepatitis c
12	259	14.2	3008	12	0933F4 hepatitis c
13	259	14.2	3010	12	068788 hepatitis c
14	259	14.2	3010	12	0901X3 hepatitis c
15	259	14.2	3010	12	0901X3 hepatitis c
16	259	14.2	3010	12	0901X3 hepatitis c

17	259	14.2	3010	12	0901X4 hepatitis c
18	259	14.2	3010	12	093077 hepatitis c
19	259	14.2	3010	12	09DTE8 hepatitis c
20	259	14.2	3010	12	09DTE6 hepatitis c
21	259	14.2	3010	12	081757 hepatitis c
22	259	14.2	3011	12	003463 hepatitis c
23	259	14.2	3013	12	0901X9 hepatitis c
24	258	14.2	3010	12	P88803 hepatitis c
25	258	14.2	3010	12	0933H8 hepatitis c
26	258	14.2	3010	12	0901Z0 hepatitis c
27	258	14.2	3010	12	09DTE1 hepatitis c
28	258	14.2	3010	12	0933H5 hepatitis c
29	258	14.2	3010	12	0933F9 hepatitis c
30	258	14.2	3010	12	0933G6 hepatitis c
31	258	14.2	3010	12	0901Y9 hepatitis c
32	258	14.2	3010	12	P90191 hepatitis c
33	258	14.2	3011	12	0933F5 hepatitis c
34	258	14.2	3013	12	0901Y0 hepatitis c
35	257	14.1	3010	12	0901Y1 hepatitis c
36	257	14.1	3010	12	09DTE9 hepatitis c
37	257	14.1	3010	12	0901Y2 hepatitis c
38	257	14.1	3010	12	0901X8 hepatitis c
39	257	14.1	3010	12	0901X7 hepatitis c
40	257	14.1	3011	12	0901X8 hepatitis c
41	257	14.1	3011	12	0901X7 hepatitis c
42	256	14.1	2284	12	081817 hepatitis c
43	256	14.1	3010	12	0901Y3 hepatitis c
44	256	14.1	3010	12	081760 hepatitis c
45	256	14.1	3010	12	P89966 hepatitis c

#### ALIGNMENTS

RESULT 1

039929 PRELIMINARY; PRT; 3008 AA.

AC 039929;  
RC 01-JAN-1998 (TREMURel. 05, Created)  
DT 01-JAN-1998 (TREMURel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMURel. 23, Last annotation update)  
DE HCV polypeptide (Genome polypeptide).  
OS Hepatitis C virus type 4a.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=31653;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=BD43 genotype 4A;  
RX MEDLINE=97335261; PubMed=9191927;  
RA Chamberlain R.W., Adams N., Saeed A.A., Simmonds P., Elliott R.M.;  
RT "Complete nucleotide sequence of a type 4 hepatitis C virus variant,  
the predominant genotype in the Middle East."  
RL J. Gen. Virol. 78:1341-1347(1997).  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
PROTEIN C AND RNA (BY SIMILARITY).  
CC EMBL; Y11604; CA72338.1; -;  
CC HSSP; P27958; 1HEI.  
CC MEROPS; S29.001; -;  
CC MEROPS; U39.001; -;  
CC InterPro; IPR001410; DEAD.  
CC InterPro; IPR002522; HCV\_capsid.  
CC InterPro; IPR002521; HCV\_core.  
CC InterPro; IPR002519; HCV\_env.  
CC InterPro; IPR002531; HCV\_NS1.  
CC InterPro; IPR002518; HCV\_NS2.  
CC InterPro; IPR004109; HCV\_NS3.  
CC InterPro; IPR000745; HCV\_NS4a.  
CC InterPro; IPR001490; HCV\_NS4b.  
CC InterPro; IPR002868; HCV\_NS5a.  
CC InterPro; IPR002166; HCV\_RdRp.

DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid.1.  
 DR Pfam: PF01542; HCV\_core.1.  
 DR Pfam: PF01539; HCV\_env.1.  
 DR Pfam: PF01560; HCV\_NS1.1.  
 DR Pfam: PF01538; HCV\_NS2.1.  
 DR Pfam: PF02907; HCV\_NS3.1.  
 DR Pfam: PF01006; HCV\_NS4a.1.  
 DR Pfam: PF01001; HCV\_NS4b.1.  
 DR Pfam: PF01506; HCV\_NS5a.1.  
 DR Pfam: PF00998; Viral\_RdRp.1.  
 DR Pfam: PF0186062; HCV\_NS1.1.  
 DR Pfam: PF0186062; HCV\_NS1.1.  
 DR SMART: SM00487; DEXDC.1.  
 DR PROSITE: PS00290; IG\_MHC.1.  
 DR PROSITE: PS00507; RDRP\_POSITIVE.1.  
 DR PROSITE: PS05021; RDRP\_VIRAL.1.  
 DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane. SEQUENCE 3008 AA; 327595 MW; 8E7FC932E27C406F CRC64;

Query Match 14.7%; Score 268; DB 12; Length 3008;  
 Best Local Similarity 31.1%; Pred. No. 1.7e-10;  
 Matches 83; Conservative 34; Mismatches 72; Indels 78; Gaps 10;

QY 23 RPAVIPDREVLVYOEFEDEMEAS-HLPYIEQGMQLAEQFOKALA--FASRQNHVSPHY 78  
 DB 1691 QPAVIPDREVLVYOEFEDEMECKHLPVHGIQLAEQFOKALGLINFGKQ----- 1742  
 QY 79 VPESDARPAIIPDREVLVHREFDEMEEFAS-RGNHVSRAHYVPESDASQAAPYIEO--- 133  
 DB 1743 --AOEATP-----VIOGNFAKLEQFMANDMMNFIISGIOYLGLSTLPGNPAIASLMSF 1793  
 QY 134 -AQVIAHQFEKYLAFASRG-----NHDSPHYVPSDAKPAI----- 170  
 DB 1794 TAAVTSPLTTOQLLNLNIGWVASQIRSDASTAVVSGALGAAGVSGLKILVDILP 1853  
 QY 171 -----IPDREVLVREFDEMEESQHL--PYIEQGMQLAE----- 201  
 DB 1854 GYGAGVAGVAVTFKIMSGMPSTEDLVNLLPALISGALVVEVVCALIRRHVGPEGAV 1913  
 QY 202 QFOKALAFASRGNHVAPTHYVTESDA 228  
 DB 1914 QMNNRLIAPASRGNHVSPHYVPSDA 1940

## RESULT 2

Q9J3G3 PRELIMINARY; PRT; 3010 AA.

DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD29;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression."  
 RL Submitted (Nov-1999) to the EMBL/GenBank/DDP databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN C AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 DR EMBL: AF207770; AAF65960.1; -  
 DR HSSP: P26663; IJXP.  
 DR InterPro: IPR000345; CytC\_heme\_bind.

DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR007045; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid.1.  
 DR Pfam: PF01542; HCV\_core.1.  
 DR Pfam: PF01539; HCV\_env.1.  
 DR Pfam: PF01560; HCV\_NS1.1.  
 DR Pfam: PF01538; HCV\_NS2.1.  
 DR Pfam: PF02907; HCV\_NS3.1.  
 DR Pfam: PF01006; HCV\_NS4a.1.  
 DR Pfam: PF01001; HCV\_NS4b.1.  
 DR Pfam: PF01506; HCV\_NS5a.1.  
 DR Pfam: PF00998; Viral\_RdRp.1.  
 DR Pfam: PF0186062; HCV\_NS1.1.  
 DR Pfam: PF0186062; HCV\_NS1.1.  
 DR SMART: SM00487; DEXDC.1.  
 DR PROSITE: PS00190; CYTOCHROME\_C.1.  
 DR PROSITE: PS00507; RDRP\_POSITIVE.1.  
 DR PROSITE: PS05021; RDRP\_VIRAL.1.  
 DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane. SEQUENCE 3010 AA; 327226 MW; 19783535650CACE3 CRC64;

Query Match 14.7%; Score 268; DB 12; Length 3010;  
 Best Local Similarity 31.9%; Pred. No. 1.7e-10;  
 Matches 102; Conservative 24; Mismatches 102; Indels 92; Gaps 13;

QY 23 RPAVIPDREVLVYOEFEDEME--ASHLPYIEQGMQLAEQFOKALAF--ASRQNHVSPHY 78  
 DB 1691 RPAVIPDREVLVREFDEMEECASHLPYIEQGMQLAEQFOKALGLITATQAQA----- 1745  
 QY 79 VPESDARPAIIPDREVLVHREFDEMEEFASR--GNHVSRAHYVPESDASQAAPYIEO--- 133  
 DB 1746 -----AAVVESKMKRL-----EAWAHMMNFIISGIOYLGLSTLPGNPAIASLMA 1792  
 QY 134 -ACVIAHQFEKYLAFASRGNHVSPHYVPSDAKPAI--IPDREVLVREFDEMEESQ 188  
 DB 1793 FTASTISPLTOSTLFLNIGWVA-AQLAPSAASAVGAGIAGAAV----- 1839  
 QY 189 HLPYIEQGMQLAEQFOKALAFASRGNHVAPTHYVTESDAKPAIYDKEVLYQYDEME 248  
 DB 1840 --GSIGLCKVLVD-----ILAGYGAGVAGALVAFKINSGERPST-----EDRDNM 1884  
 QY 249 AFASG-----NHVAPTHYVPSDASLPYMETALINQFEKYLAFASR 295  
 DB 1885 AILSPGALVGVCAIIRRHVGPEGAV-----QMNNRLIAPASR 1925  
 QY 296 GNHVSRAHYVPSDASQAAPYV 315  
 DB 1926 GNHVSRAHYVPSDASQAAPV 1945

## RESULT 3

Q81825 PRELIMINARY; PRT; 3010 AA.

ID Q81825  
 AC Q81825;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE MRNA, complete cds (Genome polyprotein).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;

Query Match	Best Local Similarity	Score	DB	Length
Matches 102; Conservative	30.7%;	268;	DB 12;	3010;
	26;	Pred. No. 1.7e-10;		
	Mismatches 89;		Indels 116;	Gaps 14;
23 RPATVPPREVLYOEFDEMEE-ASHLPYIEQGMOLAEQKOKALAF---ASRQNHVSPHY 78	1691 KPATVPPREVLYOEFDEMEEASHLPYIEQGMOLAEQKOKALGLQTRTQALAE- 1745			
79 VPESDARPAIIPDREVLRHREFDEMEEAFAFR-GNHVSPAHYVPESDASQAAPYIEQAOVI 137	1746 ----AAP-----VVESKRALETFTAKKIMNMFISIQVLAGSTLPGNPALRSP--- 1790			
138 AHQEKVLAFAFSGNHNDSPTHYVPESDAKPAIIPDREVLYREFDEMEESQH-LPYIEQG 196	1791 -----MAFTA-----SITSPL-----TTOHTLLFNILG 1813			
197 MMLAEQKOKALAFASRGNHVAPFHYVYESAK-----PALYPDREVLYOQ 242	1814 GWVAQAQAPPAAPAAAFVAGAGIAGAAGTTGIGKVLVDILAGYAGVAGALAAFK-IMSGE 1872			
243 YDEMEE-----AFASRG-----NHVAPFHYVYESDASASLPYMDETRAIAG 283	1873 MPASMEDVNNLLPALLSPGALVVGICVCAAILRRHGPGBGAV----- 1913			
284 QFEKVLAFASRGNHVSPHYVPESEPOVVT 315	1914 QMMNRLIAFASRGNHVSPHYVPESEPARAVT 1945			

ID	O68285		PRELIMINARY;	PRT;   3010 AA.
AC	O68285:			
DT	01-NOV-1996	(TReMBLeRel_01,	Created)	
DT	01-NOV-1996	(TReMBLeRel_01,	Last sequence update)	
DT	01-MAR-2003	(TReMBLeRel_23,	Last annotation update)	
DE	Genome polypeptidein.			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
CC	Hepadnavirus			
CX	NCBI_Taxid=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Cho J., Park Y., Lee Y., Yang J., Ryu W.;			
RT	"Molecular cloning of Hepatitis C Virus genome from chronic patients in Korea."			
RL	Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.			
CC	-1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).			
CC	EMBL U16362; AAA52748.1; -.			
DR	HSP, p26663; 1TXP.			
DR	InferPro: IPR001410; DEAD.			
DR	InferPro: IPR002522; HCV_capsid.			
DR	InferPro: IPR002521; HCV_core.			
DR	InferPro: IPR002519; HCV_env.			
DR	InferPro: IPR002531; HCV_NS1.			
DR	InferPro: IPR002518; HCV_NS2.			
DR	InferPro: IPR004109; HCV_NS3.			
DR	InferPro: IPR000745; HCV_NS4A.			
DR	InferPro: IPR001490; HCV_NS4B.			
DR	InferPro: IPR002868; HCV_NS5A.			
DR	InferPro: IPR002165; HCV_RdRp.			
DR	InferPro: IPR007095; RNA_pol_DS_PS.			
DR	InferPro: IPR007094; RNA_pol_PSVIrf.			
DR	Pfam: PF01543; HCV_capsid.1.			
DR	Pfam: PF01542; HCV_core.1.			
DR	Pfam: PF01539; HCV_env.1.			
DR	Pfam: PF01560; HCV_NS1.1.			
DR	Pfam: PF01538; HCV_NS2.1.			
DR	Pfam: PF02907; HCV_NS3.1.			
DR	Pfam: PFO1006; HCV_NS4A.1.			
DR	Pfam: PFO1001; HCV_NS4B.1.			
DR	Pfam: PF01506; HCV_NS5A.1.			
DR	Pfam: PFO0998; Viral RdRp.1.			
DR	proDom; PD186062; HCV_NS1.1.			
DR	SMART; SMO0487; DEXdc.1.			
DR	PROSITE; PSS0507; RDRP_POSITIVE.1.			
DR	PROSITE; PSS0521; RDRP_VIRAL.1.			
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane. SEQUENCE     3010 AA;   326914 MW;   298306FCSB8EBC9E CRC64;			
SQ				
Query Match	14.7%	Score 268;	DB 12;	Length 3010;
Best Local Similarity	30.7%;	Pred. No. 1,7e-10;		
Matches 102;	Conservative 26;	Mismatches 88;	Indels 116;	Gaps 14;
DY	23 RPAAVIPREVLTYGOEPDMEF-AASHLPITBOGOMLAEFKOKALAFA---ASRGNHVSPTHY 78	: ::   ::		
Db	1691 KPAPIIIPREVLYOEFEDEMESCASHLPYEFGQMOLAEFKOKAGLLGTATKOABA----- 1745	:: :: ::		
DY	79 VPESDAARPPIIDPREVLHREFDEMEAFAFR-GNHVPAYHVPPESDAQAPYLEQAQVI 137	:: :: ::		
Db	1746 -----AAP-----VESKMKALEFTWMGHMMNFISGIQLAGLSLTPNGPAISRPF-- 17900	:		
DY	138 AHOFKEKVLFAPSARGNHDSPTHYVPESDAKPAPIIDPREVLRYREFDEMESOH-LPYIEOG 196	:: :: ::		
Db	1791 -----MAFTA-----SITSPL-----TTQHTLFLFNIG 1813	:		
DY	197 MLMLAEQFKOKALAFPSRGHNVARPTHVTSDAK-----PALVPDKEVLYOO 242	:: :: ::		
Db	1814 GWAADAOAPPASAASFVGAGTAGAACVGTICGLKLVLDILAGYGAVAGALTVAER-TMSGE 18722	:		

QY 243 YDEME-----AFASRG-----NHVAPTHYVESDASASLPYMETRAIAG 283  
 DB 1873 MPSEEDVNLPLPALSGALVGVCAIIRRHVPEEGAV-----1913  
 QY 284 QFKEKVLAFASRGNHVSPRHYVESEPOVVT 315  
 DB 1914 QMMNRLIAFASRGNHVSPRHYVESEPAARVT 1945  
 RESULT 5  
 ID 090IX1 PRELIMINARY; PRT; 3010 AA.  
 AC 090IX1;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NCBI\_TaxID=11103;  
 [1]  
 K. SEQUENCE FROM N.A.  
 RC STRAIN=MD10-2;  
 RX MEDLINE=20013325; PubMed=10544098;  
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,  
 RA Tazawa J.I., Izumi N., Marumo F., Sato C.;  
 RT "Time-related changes in full-length hepatitis C virus and hepatitis  
 RT activity.";  
 RL Virology 263:244-253(1999).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD10-2;  
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,  
 RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;  
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL: AF165064; AAD56199.1; -;  
 CC HSP: P26663; IJXP.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid.1.  
 DR Pfam: PF01542; HCV\_core.1.  
 DR Pfam: PF01539; HCV\_env.1.  
 DR Pfam: PF01560; HCV\_NS1.1.  
 DR Pfam: PF01538; HCV\_NS2.1.  
 DR Pfam: PF02907; HCV\_NS3.1.  
 DR Pfam: PF01006; HCV\_NS4a.1.  
 DR Pfam: PF01001; HCV\_NS4b.1.  
 DR Pfam: PF01506; HCV\_NS5a.1.  
 DR Pfam: PF00998; Viral\_RdRp.1.  
 DR Pfam: PF01538; HCV\_NS1.1.  
 DR SMART: SM00487; DEXDC.1.  
 DR PROSITE: PSS0507; RDRP\_POSITIVE.1.  
 DR PROSITE: PSS0521; RDRP\_VIRAL.1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SO SEQUENCE 3010 AA; 326902 MW; 0423385981E71EDD CMC64;

Query Match 14.5%; Score 263; DB 12; Length 3010;  
 Best Local Similarity 29.7%; Pred. No. 3.9e-10;  
 Matches 94; Conservative 30; Mismatches 107; Indels 86; Gaps 11;  
 QY 23 RPAPIIDREVLVYOEPEDEME-ASHLPYIEQKQALAEQFOKXALAF---ASRONHVSPTHY 78  
 DB 1691 RPAIIPDREVLVREFPEMEECASHLPYIQGQALAEQFOKXALGLLQITKQ-----1742  
 QY 79 VPESDARPAIIPDREVLVAREFDEMEEFASR-GNHVSPAHYVPESDASQAAPYIEQAQVI 137  
 DB 1743 -----AEAAVP-----VVEESKQALAEAFMAKHMNFISGIQYLAGSTLPGNALI-----1787  
 QY 138 ANQFKEKVLAFASRGNHVSPRHYVESEPOVVT 315  
 DB 1788 -----ASLMAFTASVTSPLTQYT-----LLFNILGQVAAQIAPPSASSA 1828  
 QY 198 MIAEQFKQALAFASRGNHVAPTHYVTESDAKPALVPDEKVLVYQYDEME-----AFA 251  
 DB 1829 FVGAGIAGAAVSGISGLKVLVILACYGAGVAGALVAFK-VMSGEVPTEDLVNLLPALI 1887  
 QY 252 SRG-----NHVAPTHYVESDASASLPYMETRAIACQFKEKVLAFASRGNH 298  
 DB 1888 SPGALVGVCAIIRRHVPEEGAV-----QMMNRLIAFASRGNH 1928  
 QY 299 VSPRHYVESEPOVVT 315  
 DB 1929 VSPRHYVESEPAARVT 1945  
 RESULT 6  
 ID 068949 PRELIMINARY; PRT; 3010 AA.  
 AC 068949;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE UK1-full protein (Genome polyprotein).  
 GN UK1-FUL.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=93013055; PubMed=1327977;  
 RA Honda M., Kaneko S., Unoura U., Kobayashi K., Murakami S.;  
 RT "Sequence comparisons for a hepatitis C virus genome RNA isolated from  
 RT a patient with liver cirrhosis.";  
 RL Gene 120:317-318(1992).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL: X61596; CAA43793.1; -;  
 CC HSP: P26663; IJXP.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid.1.  
 DR Pfam: PF01542; HCV\_core.1.  
 DR Pfam: PF01539; HCV\_env.1.  
 DR Pfam: PF01560; HCV\_NS1.1.  
 DR Pfam: PF01538; HCV\_NS2.1.

```

DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1
FT CHAIN 191
FT CHAIN 383
FT CHAIN 384
FT CHAIN 734
FT CHAIN 1007
FT CHAIN 1616
FT CHAIN 1862
FT CHAIN 2013
FT CHAIN 2014
SEQUENCE 3010 AA; 327178 MW; 4517408FF613ADA CRC64;

Query Match 14.5%; Score 263; DB 12; Length 3010;
Best Local Similarity 32.5%; Pred. No. 3,9e-10;
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 11;

QY 23 RPAVIPDREVLVYOEEDMEEE-ASHLPYIEQGMQLAEQFOKALAF---ASRGNHVSPTHY 78
DB 1691 RPAIIPDREVLVYOEEDMEECASHLPYIEQGMQLAEQFOKALGLLOTATSKDAEA----- 1745

QY 79 VPESDAPPAIIPDREVLVHREFDEMEEAFAFR--GNHVSAPHYVPESDASQAAPYIEQ--- 133
DB 1746 -----AAPVESKMWQL-----EAFWAKHMMNFIISGIQYLAGLSTLPGNPAIVSLMA 1792

QY 134 --AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB 1793 FTASITSPLTQTHTLLENIILGWAQAQLAPPSAASAFVAGIAGAAGSIGLKVLDIL 1852

QY 171 -----IPDREVLVREFD-EMESQHL-----PYIEQGMQL-----A 200
DB 1853 AGYGAGVAGALVAFKMGSEMPSTEDLVNLLPALISPGALVGVCAILRRHVGPEGGA 1912

QY 201 EQFOKALAFASRGNHVAPTHYTESDA 228
DB 1913 VQMMNRLIAFASRGNHVSPTHYVPESDA 1940

RESULT 7
O9JTE2 PRELIMINARY; PRT; 3010 AA.
AC O9JTE2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN NCBI_Taxid=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=HCV1169;
RA Takehashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited.";
RT Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC EMBL; AB049095; BAB18808.1; -
CC DR HSPF; P26663; IJXP.

```

```

DR InterPro; IPR000345; Cyto_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01542; HCV_capsid; 1.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SEQUENCE 3010 AA; 326867 MW; 25BE54B9D7EEA15 CRC64;

Query Match 14.4%; Score 261; DB 12; Length 3010;
Best Local Similarity 32.1%; Pred. No. 5.5e-10;
Matches 86; Conservative 28; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVIPDREVLVYOEEDMEEE-ASHLPYIEQGMQLAEQFOKALAF---ASRGNHVSPTHY 78
DB 1691 RPAIIPDREVLVYOEEDMEECASHLPYIEQGMQLAEQFOKALGLLOTATQAEV----- 1745

QY 79 VPESDAPPAIIPDREVLVHREFDEMEEAFAFR--GNHVSAPHYVPESDASQAAPYIEQ--- 133
DB 1746 -----AAPVESKMRAL-----EAFWAKHMMNFIISGIQYLAGLSTLPGNPAIVSLMA 1792

QY 134 --AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB 1793 FTASITSPLTQTHTLLENIILGWAQAQLAPPSAASAFVAGIAGAAGSIGLKVLDIL 1852

QY 171 -----IPDREVLVREFD-EMESQHL-----PYIEQGMQL-----A 200
DB 1853 AGYGAGVAGALVAFKMGSEMPSTEDLVNLLPALISPGALVGVCAILRRHVGPEGGA 1912

QY 201 EQFOKALAFASRGNHVAPTHYTESDA 228
DB 1913 VQMMNRLIAFASRGNHVSPTHYVPESDA 1940

RESULT 8
O9J3H9 PRELIMINARY; PRT; 3010 AA.
ID O9J3H9
AC O9J3H9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN NCBI_Taxid=11103;
RP SEQUENCE FROM N.A.
RC STRAIN=MD13;

```



Query Match	14.4%	Score 261	DB 12	Length 3010
Best Local Similarity	32.6%	Pred. No. 5,5e-10		
Matches 87	Conservative 27	Mismatches 75	Indels 78	Gaps 10
QY 23	RPAYIPDREVLVYQEPDEMEF-AASHPIYIEQGMQLAEQKOKALAF--ASRONHVSPTHY 78			
	1691	RPAYIPDREVLVYREFDEMEECASHPIYIEQGMQLAEQKOKALAGLLOATATKQAEA----- 1745		
QY 79	VPESDARPAIIPDREVLVHREFDEMEAFASR-GNHVSPAHVVPESDAQAPYIEQ----- 133			
	1746	-----AAPAV-----ESKRALETFWAKKMMNFISGIQYIAGISTLPGNPAIASLMAF 1793		
QY 134	-AOVIAHOFKEKVLAF-----ASRGNDSPTHVVPESDAKPAI----- 170			
QY 1794	TASVTSPLTTQNTLLFNILGGMVAQLARPPSAASAFVAGIAGAAGVSGIGLKVLDVILA 1855			
QY 171	-----IDREVLVYREFDEMEESQHL--PYIEQGMML-----AE 201			
QY 1854	GYGAGVACALAFKMGSETSPSAEVLVLLPALISPGALVVGVCALIRRHVGPGEVAG 1913			
QY 202	QFKOKALAFASRGNHVAPTHVVTESDA 228			
QY 1914	QMMNRLIAFASRGNHVSPTHVVPESDA 1940			
RESULT 9				
090IX2				
090IX2	PRELIMINARY:	PRT:	3010 AA.	
090IX2				
01-MAY-2000	(TRENDArel. 13, Created)			

DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
OS	Genome polyprotein.
OS	Hepatitis C virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepadnavirus.
NC	NCBI_Taxid=11103;
RN	(1)
RP	SEQUENCE FROM N.A.
RC	STRAIN=MD10-1;
RA	MEDLINE=20013325, PubMed=10544098;
RL	Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
CC	Tazawa J., Izumi N., Matsumo F., Sato C.;
CC	"Time-related changes in full-length hepatitis C virus and hepatitis
CC	activity.";
CC	Virology 263:244-253(1999).
CC	[2]
CC	SEQUENCE FROM N.A.
CC	STRAIN=MD10-1;
CC	Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
CC	Sakamoto N., Fukuma T., Tazawa J., Izumi N., Matsumo F., Sato C.;
CC	Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC	LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC	PROTEIN C AND MNA (BY SIMILARITY).
CC	EMBL; AF165063; AAD56198.1; -
CC	HSSP; P26663; 1UXP.
CC	InterPro: IPR001410; DEAD.
CC	InterPro: IPR002522; HCV_capsid.
CC	InterPro: IPR002521; HCV_core.
CC	InterPro: IPR002519; HCV_env.
CC	InterPro: IPR002531; HCV_NS1.
CC	InterPro: IPR002518; HCV_NS2.
CC	InterPro: IPR004109; HCV_NS3.
CC	InterPro: IPR000745; HCV_NS4.
CC	InterPro: IPR001490; HCV_NS4b.
CC	InterPro: IPR002868; HCV_NS5a.
CC	InterPro: IPR002166; HCV_RdRp.
CC	InterPro: IPR007095; RNA_pol_DS_PS.
CC	InterPro: IPR007094; RNA_pol_LSVir.
CC	Pfam; PF01543; HCV_capsid; 1.
CC	Pfam; PF01542; HCV_core; 1.
CC	Pfam; PF01539; HCV_env; 1.
CC	Pfam; PF01560; HCV_NS1; 1.
CC	Pfam; PF01538; HCV_NS2; 1.
CC	Pfam; PF02907; HCV_NS3; 1.
CC	Pfam; PF01006; HCV_NS4a; 1.
CC	Pfam; PF01001; HCV_NS4b; 1.
CC	Pfam; PF01506; HCV_NS5a; 1.
CC	Pfam; PF00998; Viral_RdRp; 1.
CC	ProDom; PD166062; HCV_NS1; 1.
CC	SMART; SM00467; DEDCP; 1.
CC	PROSITE; PS50507; RDRP_POSITIVE; 1.
CC	PROSITE; PS50521; RDRP_VIRAL; 1.
CC	Coat protein, Envelope protein; Glycoprotein; Nonstructural protein;
CC	KM Polyprotein; RNA-directed RNA polymerase; Transmembrane.
CC	SEQUENCE 3010 AA; 326855 MW; 38064AFA819ED552 CRC64;
CC	SO
CC	Query Match 14.4%; Score 261; DB 12; Length 3010;
CC	Best Local Similarity 29.9%; Pred. No. 5.5e-10;
CC	Matches 95; Conservative % 26; Mismatches 107; Indels 88; Gaps 11
CC	23 RPAVIPPREVLYQGFDEMEF-ASHLPYIEGCMQLAEOFKQALAF---ASQNHVSPHY 78
CC	DB 1691 RPAIIPPREVLYRFDEMECASHLPYIEGCMQLAEOFKQALLOTATQAE----- 1745
CC	OY 79 VPESDAPPAIIPREVLYHREFDEMEFAFSR--GNHVSFAHYVPESDASQAPYEQOV 136
CC	DB 1746 -----AAPVESKQAL-----EAFNAKHMNFISGIOLYLAGLSTLPENPAI----- 1787
CC	OY 137 IAHOFKEKVLAFASRCGNHDSPTHYVPSDAKPAIIPREVLYRFDEMEESQHLPIEEOG 196

```

DB 1788 -----ASLMFTASVTSPITTYT-----LLENILGCVWVAQIAPPSASS 1827
QY 197 MMLAEQFKOKALAFASRGHNVAPTHVETESDAKPAVPDKEVLVYQYDEME-----AF 250
DB 1828 AFGVGAGIAGAAVGSIGLGVLDIILAGYAGVAGALVAFK-VMSGEVSTELVLLPAIL 1886
QY 251 ASRG-----NHVAPTHVVESDASASLPYDETRAIAGQFKVLAFAASRCN 297
DB 1887 LSPGALVGVCAAILRRHVGPGEAV-----QNMNRLIAFASRCN 1927
QY 298 HVSFPHVPESEPOVVT 315
DB 1928 HVSPTHVPESDAAARVT 1945

RESULT 10
ID 093016 PRELIMINARY; PRT; 3010 AA.
AC 093016;
I 01-NOV-1998 (TREMBLrel. 08, Created)
I 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-A;
RA Trowbridge R.; Gowans E.J.;
RL Submitted (Jul-1997) to the EMBL/Genbank/DBJ databases.
RM [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-A;
RX MEDLINE=9823263; Pubmed=9572551;
RA Trowbridge R.; Gowans E.J.;
RT "Molecular cloning of an Australian isolate of hepatitis C virus.";
RL Arch. Virol. 143:501-511(1998)
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC EMBL; AJ000009; CAA03854.1; -.
DR HSP; P26663; IXP.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVIT.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01539; HCV_core; 1.
DR Pfam; PF01560; HCV_env; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD016062; HCV_NS1; 1.
DR SMART; SMD0487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KV Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

```

```

KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191 CORE PROTEIN.
FT CHAIN 192 383 E1 PROTEIN.
FT CHAIN 384 746 E2 PROTEIN.
FT CHAIN 747 809 P7 PROTEIN.
FT CHAIN 810 1026 NS2 PROTEIN.
FT CHAIN 1027 1657 NS3 PROTEIN.
FT CHAIN 1658 1711 NS4A PROTEIN.
FT CHAIN 1712 1972 NS4B PROTEIN.
FT CHAIN 1973 2419 NS5A PROTEIN.
FT CHAIN 2420 3010 NS5B PROTEIN.
SQ SEQUENCE 3010 AA; 327120 MW; 0E02EDA54A861D CRC64;

Query Match 14.3%; Score 260; DB 12; Length 3010;
Best Local Similarity 32.6%; Pred. No. 6,4e-10;
Matches 87; Conservative 29; Mismatches 73; Indels 78; Gaps 10;

QY 23 RPAVIPDEVLVYQFDEME--ASHLPYIEQGMQLAEQFKOKALAF---ASRQNHVSPHY 78
DB 1691 RPAIIPDEVLVYQFDEMEECASHLPYIEQGMQLAEQFKOKALGILQTRTKQ----- 1742
QY 79 VPESDARPAIIPDEVLHREPDMEBAFASR-GHVSFAHYVP-----ES 122
DB 1743 -----AEAAVP---VVESKWQALEAFWAKMMNFISGIQYLAGLSTLPGNLAIASLMAF 1793
QY 123 DASQAPYIEQAOVI-----AHQFKVLAFAASRGHNSPPTYVPESD 165
DB 1794 TASIITPITHTHTLFLNLGCVWVAQIAPPSAASAFVGAGIAGAAVGSIGLGVLDILA 1853
QY 166 AKPAIIPDEVLVYREFD-EMESQHL-----PYIEQGMQL-----AE 201
DB 1854 GYGAGVAGALVAFKMGSEMGSTEDLVNLLPAILSPGALVGVCAAILRRHVGPGEAV 1913
DB 1914 QNMNRLIAFASRGHNVSPHYVPESDA 1940

QY 202 QFKOKALAFASRGHNVAPTHVETESDA 228
DB 1914 QNMNRLIAFASRGHNVSPHYVPESDA 1940

RESULT 11
ID 093037 PRELIMINARY; PRT; 3010 AA.
AC 093037;
I 01-OCT-2000 (TREMBLrel. 15, Created)
I 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD25;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL; AF207766; AAF65956.1; -.
DR HSP; P26663; IXP.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.

```

DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR PROSITE: PS50507; RDRP\_POSITIVE; 1.  
 DR PROSITE: PS50521; RDRP\_VIRAL; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 K<sub>m</sub> RNA-directed RNA polymerase; Transferase; Transmembrane.  
 K<sub>m</sub> RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326507 MW; 9EBACAD4B441DF CRC64;

Query Match 14.3%; Score 260; DB 12; Length 3010;  
 Best Local Similarity 32.1%; Pred. No. 6.4e-10;  
 Matches 86; Conservative 27; Mismatches 75; Indels 80; Gaps 10;  
 QY 23 RPVAVDREVLVYOEFDMEEF-ASHLPYIEQGQQLAEQFKOKALAF---ASRGNHVSPTHY 78  
 DB 1691 RPVAVDREVLVYOEFDMEEFASHLPYIEQGQQLAEQFKOKALGLGLQTATKQAEA----- 1745  
 QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSFAHYVPESDASQAAPYIEQ--- 133  
 DB 1746 ----AAPVESKRAL-----DAFMKMMWNTISGIQYLAGISTLPGNPAIASLMA 1792  
 QY 134--AQVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170  
 DB 1793 FTASITSPLTQSTLTFNIGWVAQAQLAPSAASAFVAGAGIAGAVSGIGKVLVDIL 1852  
 QY 171 -----IPREVLVYREFDMEESQHL--PYIEQGQML-----A 200  
 DB 1853 AGYAGVAGALVAFKMGSEMPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGEA 1912  
 QY 201 EOKKALAFASRGNHVAPTHYTESDA 228  
 DB 1913 VQMMNRLLIAFASRGNHVSPTHYVPESDA 1940

RESULT 12  
 Q9DTES PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTES; 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI Taxid=11103;  
 RN [1]\_  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV145;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 with hepatocellular carcinoma: the 'progression score' revisited";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 DR EMBL: AB049092; BAB18805.1; -  
 DR HSSP: P27958; 1HEF.  
 DR InterPro: IPR000345; Cyt\_c\_heme\_bind.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002511; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 2.  
 DR PROSITE: PS50507; RDRP\_POSITIVE; 1.  
 DR PROSITE: PS50521; RDRP\_VIRAL; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 K<sub>m</sub> Hydrolyase; Nonstructural protein; Polypeptide;  
 K<sub>m</sub> RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326986 MW; 573C0F3C55B3F3F4 CRC64;

Query Match 14.3%; Score 260; DB 12; Length 3010;  
 Best Local Similarity 31.8%; Pred. No. 6.4e-10;  
 Matches 85; Conservative 29; Mismatches 75; Indels 78; Gaps 10;  
 QY 23 RPVAVDREVLVYOEFDMEEF-ASHLPYIEQGQQLAEQFKOKALAF---ASRGNHVSPTHY 78  
 DB 1691 RPVAVDREVLVYOEFDMEEFASHLPYIEQGQQLAEQFKOKALGLGLQTATKQAEA----- 1745  
 QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSFAHYVPESDASQAAPYIEQ--- 133  
 DB 1746 ----AAP-----VSESKRALGFMKMMWNTISGIQYLAGISTLPGNPAIASLMAF 1793  
 QY 134--AQVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170  
 DB 1794 TASITSPLTQSTLTFNIGWVAQAQLAPSAASAFVAGAGIAGAVSGIGKVLVDILA 1853  
 QY 171 -----IPREVLVYREFDMEESQHL--PYIEQGQML-----AE 201  
 DB 1854 GYAGVAGALVAFKMGSEMPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGEAV 1913  
 QY 202 OFKOKALAFASRGNHVAPTHYTESDA 228  
 DB 1914 QMMNRLLIAFASRGNHVSPTHYVPESDA 1940

RESULT 13  
 Q9J3F4 PRELIMINARY; PRT; 3008 AA.  
 AC Q9J3F4; 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)



```

Db      1746 -----AAP-----VSESKRALFTFWAKHMMNFISGIQVLAGSLTLPGNPAI----- 1787
QY      138 AHQREKVLAFASRGNDSPHYVPESDAKPAIIPDREVLYREFDEMEESCHLPYIEGGM 197
Db      1788 -----ASMAFTASTISPLTQYT-----LLENIIGCWVAQAQAPPSAASA 1828
QY      198 MLAEOFKOKALAFASRGNHVAPTHVTESDAKPALVPKEXVLYQOYDME-----AFA 251
Db      1829 FVGAGIAGAAVGSIGLKVLDIILAGYAGVAGALVAFK-VMSGDMSTEDLVNLLPATL 1887
QY      252 SRG-----NHVAPTHVYESDASASLPYMETRAIAGQFEKVLAFASRGNH 298
Db      1888 SPGALVGVCAAILRRHVGPGEAGV-----QMMNRLIAFASRGNH 1928
QY      299 VSPRHYVPESEPQVVT 315
Db      1929 VSPTHVYVESDAAARVT 1945

```

## JUL 15

```

:K3
:Q90IX3 PRELIMINARY; PRT; 3010 AA.

```

```

AC      Q90IX3;
DT      01-MAY-2000 (Tremblrel. 13, Created)
DT      01-MAY-2000 (Tremblrel. 13, last sequence update)
DT      01-MAR-2003 (Tremblrel. 23, last annotation update)
DE      Genome polypeptide.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_Taxid=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=MD9-2;
RX      MEDLINE=20013325; PubMed=10544098;
RA      Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
RA      Tazawa J., Izumi N., Marumo F., Sato C.;
RT      "time-related changes in full-length hepatitis C virus and hepatitis
RT      activity.";
RL      Virology 263:244-253 (1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MD9-2;
RA      Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
RA      Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RL      Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC      -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
        LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
        PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
        PROTEIN C AND MRNA (BY SIMILARITY).

```

```

DR      EMBL; AF165062; AAD56197.1; -.
DR      HSP; P26663; INS3.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR002522; HCV_capsid.
DR      InterPro; IPR002521; HCV_core.
DR      InterPro; IPR002519; HCV_env.
DR      InterPro; IPR002531; HCV_NS1.
DR      InterPro; IPR002518; HCV_NS2.
DR      InterPro; IPR004109; HCV_NS3.
DR      InterPro; IPR000745; HCV_NS4a.
DR      InterPro; IPR001490; HCV_NS4b.
DR      InterPro; IPR002868; HCV_NS5a.
DR      InterPro; IPR002166; HCV_NS5b.
DR      InterPro; IPR007095; RNA_pol_DS_PS.
DR      InterPro; IPR007094; RNA_pol_PSVir.
DR      Pfam; PF01543; HCV_capsid; 1.
DR      Pfam; PF01542; HCV_core; 1.
DR      Pfam; PF01539; HCV_env; 1.
DR      Pfam; PF01560; HCV_NS1; 1.
DR      Pfam; PF01538; HCV_NS2; 1.
DR      Pfam; PF02907; HCV_NS3; 1.
DR      Pfam; PF01006; HCV_NS4a; 1.

```

```

DR      Pfam; PF01001; HCV_NS4b; 1.
DR      Pfam; PF01506; HCV_NS5a; 1.
DR      Pfam; PF00998; Viral_RdRp; 1.
DR      ProDom; PD166062; HCV_NS1; 1.
DR      SMART; SM00487; DEXDC; 1.
DR      PROSITE; PSS0507; RdRp_POSITIVE; 1.
DR      PROSITE; PSS0521; RdRp_VIRAL; 1.
KW      Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW      Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ      SEQUENCE 3010 AA; 327253 MW; 9F1B0B3F536774FA CRC64;

```

```

Query Match      14.2%; Score 259; DB 12; Length 3010;
Best Local Similarity 31.5%; Pred. No. 7,6e-10;
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;

```

```

QY      23 RPAVIPDREVLYOEFDEMEE-ASHLPYIEGQMLAEOFKOKALAF---ASRGNHVSPHY 78
Db      1691 RPAVIPDREVLYREFDEMEESCHLPYIEGQMLAEOFKOKALGLLOTATKQ----- 1742
QY      79 VPESDARPAIIPDREVLYREFDEMEESAFSR-GNHVSPHYVPESDASQAAPYIEQ--- 133
Db      1743 -----AEAAPV---VSESKRALFTFWAKHMMNFISGIQVLAGSLTLPGNPAIASLMAF 1793
QY      134 -AQVIAHQREKVLAF-----ASRGNDSPHYVPESDAKPAI----- 170
Db      1794 TASITSPLTQNTLLENIIGCWVAQAQAPPSAASAFVAGIAGAAVGSIGLKVLDIILA 1853
QY      171 -----IPDREVLYREFDEMEESCHL--PYIEGQML-----AE 201
Db      1854 GYGAGVAGALVAFKIMGEMPTEDLVNLLPATLSPGALVGVCAAILRRHVGPGEAGV 1913
QY      202 QFKOKALAFASRGNHVAPTHVYESDA 228
Db      1914 QMMNRLIAFASRGNHVSPHYVPESDA 1940

```

Search completed: July 25, 2003, 13:11:59.  
Job time : 100 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: July 25, 2003, 12:55:06 ; Search time 23 Seconds

(without alignments)  
719.713 Million cell updates/sec

Title: US-09-491-146a-52  
Perfect score: 1818  
Sequence: 1 AAHPIYLEQGHLEQFKQK.....KALIEQGMALMKSIQ 352

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

1 number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	260	14.3	3010	1	P26662 h genome po
2	258	14.2	3010	1	P26663 h genome po
3	256	14.1	3010	1	P26664 h genome po
4	254	14.0	3010	1	P26665 h genome po
5	254	14.0	3011	1	P26666 h genome po
6	247	13.6	3011	1	P26667 h genome po
7	203.5	11.2	3033	1	P26668 h genome po
8	201.5	11.1	3033	1	P26669 h genome po
9	121.5	6.7	911	1	P26670 h genome po
10	120.5	6.6	5327	1	P26671 h genome po
11	114.5	6.3	912	1	P26672 h genome po
12	114.5	6.3	4473	1	P26673 h genome po
13	113.5	6.2	911	1	P26674 h genome po
14	113.5	6.2	4684	1	P26675 h genome po
15	109.5	6.0	4687	1	P26676 h genome po
16	107.5	5.9	1733	1	P26677 h genome po
17	107	5.9	910	1	P26678 h genome po
18	106.5	5.9	1736	1	P26679 h genome po
19	106	5.8	2748	1	P26680 h genome po
20	104	5.7	964	1	P26681 h genome po
21	103	5.7	3259	1	P26682 h genome po
22	103	5.7	5430	1	P26683 h genome po
23	103	5.7	5938	1	P26684 h genome po
24	101.5	5.6	788	1	P26685 h genome po
25	101.5	5.6	2517	1	P26686 h genome po
26	101	5.6	478	1	P26687 h genome po
27	101	5.6	1036	1	P26688 h genome po
28	101	5.6	3678	1	P26689 h genome po
29	100	5.5	478	1	P26690 h genome po
30	100	5.5	649	1	P26691 h genome po
31	100	5.5	1036	1	P26692 h genome po
32	100	5.5	1036	1	P26693 h genome po
33	100	5.5	1745	1	P26694 h genome po

34	99	5.4	631	1	SYR_HAUNI	Q9huz2 halobacteri
35	98.5	5.4	904	1	AAQ4_CHICK	Q90734 gallus gall
36	98.5	5.4	932	1	IF3A_SCHPO	O74760 schizosacch
37	97.5	5.4	468	1	SYE_THERM	P27000 thermus the
38	97.5	5.4	1026	1	YH2O_CAEEL	Q23223 caenorhabdi
39	97	5.3	632	1	CCB2_RABIT	P54288 oryctolagus
40	97	5.3	660	1	CCB2_HUMAN	O08289 homo sapien
41	97	5.3	1756	1	PEPL_HUMAN	O60437 homo sapien
42	96.5	5.3	431	1	RAPH_BACSU	P40771 bacillus su
43	96.5	5.3	562	1	ILVD_STRAM	Q99839 staphylococ
44	96	5.3	2805	1	MAPA_HUMAN	P78559 homo sapien
45	96	5.3	5120	1	PCLO_CHICK	Q9p36 gallus gall

## ALIGNMENTS

RESULT 1  
POLG\_HCVJA STANDARD; PRT; 3010 AA.  
AC P26662; 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22);  
DE Envelope glycoprotein E1 (GP12) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadivirin)  
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate Japanese) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_Taxid=11116;  
RN [1]  
RP MEDLINE=9108550; PubMed=2175903;  
RX Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
RA Sugimura T., Shimotohno K.;  
RT Japanese cloning of the human hepatitis C virus genome from  
Proc. Natl. Acad. Sci. U.S.A. 87:9524-9526(1990).  
[2]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=91192160; PubMed=1849488;  
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,  
OHkoshi S., Shimotohno K.;  
RT "Molecular structure of the Japanese hepatitis C viral genome.";  
FEBS Lett. 280:325-328(1991)  
RL  
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polypeptide, commonly with Asp or Glu in the P6  
CC position. Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +  
CC (RNA) (N).  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D90208; BAA14233.1; -

DR PIR; A39253; GNMVCJ.  
 DR HSP; P26663; IUXP.  
 DR MEROPS; S29.001; -.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR004109; HCV\_NS3.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_NS5b.  
 DR InterPro; IPR001650; Helicase C.  
 DR InterPro; IPR007095; RNA pol Ds.PS.  
 DR InterPro; IPR007094; RNA\_pol\_Pstir.  
 DR PIR; P01543; HCV\_capsid; 1.  
 DR PIR; P01542; HCV\_core; 1.  
 DR PIR; P01539; HCV\_env; 1.  
 DR PIR; P01560; HCV\_NS1; 1.  
 DR PIR; P01538; HCV\_NS2; 1.  
 DR PIR; P01597; HCV\_NS3; 1.  
 DR PIR; P01006; HCV\_NS4a; 1.  
 DR PIR; P01001; HCV\_NS4b; 1.  
 DR PIR; P01506; HCV\_NS5a; 1.  
 DR PIR; P00271; helicase C; 1.  
 DR PIR; P00998; Viral RdRp; 1.  
 DR PIR; P01862; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXdc; 1.  
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 DR Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.  
 DR INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT ACT\_SITE 1230 1237  
 FT NP\_BIND 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2788 2788  
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Matches 86; Conservative 28; Mismatches 75; Indels 78; Gaps 10;  
 QY 23 RPAVIPDEVLVYQFEDMEE-ASHLPYIEQGMQLAEQFKALAF--ASRGNHVSPTHY 78  
 DB 1691 RPAVIPDEVLVYQFEDMEECASHLPYIEQGMQLAEQFKALGILQATKAEA----- 1745  
 QY 79 VPESDARPAIIPDEVLVIREFDEMEARSR-GNHVSPAHVYVESDASQAPYIQO---- 133  
 DB 1746 ----AAP-----VVESKRALVFWAKHMMNFIISGIQVLGSLTPGNPAIASLMAF 1793  
 QY 134 -AQVIAHQFKEKVLAF-----ASRGNHVSPTHYVESDAPKPAI----- 170  
 DB 1794 TASITSPITQTQTLFNLIGKVAQAQLAPPSAASFAGAGIAGAAVSGISGLKVLVDILA 1853  
 QY 171 -----IPDEVLVIREFDEMEESQHL--PYIEQGMQLAEQFKALAF-----AE 201  
 DB 1854 GYGAGVAGALVAFKXMGSEMPSTEDLVLLPAILSPALVGVCAAILRRHVGGEGAV 1913  
 QY 202 QFKQALAFASRGHVAAPTHYTESDA 228  
 DB 1914 QMMNRLIAFASRGHVSPTHYVESDA 1940  
 RESULT 2  
 POLG\_HCVTW  
 ID POLG\_HCVTW STANDARD; PRT; 3010 AA.  
 AC P29846;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein (contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepaticityin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)}.  
 OS Hepatitis C virus (isolate Taiwan) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxId=31645;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230206; PubMed=1314449;  
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
 RT "The Taiwanese hepatitis C virus genome: sequence determination and  
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";  
 RL Virology 188:102-113(1992).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M84754; -. NOT\_ANNOTATED\_CDS.  
 DR PIR; A40244; GNMVTV.  
 DR PDB; 1N64; 25-FEB-03.

Query Match 14.3%; Score 260; DB 1; Length 3010;  
 Best Local Similarity 32.2%; Pred. No. 4,1e-09;

```

Query March 14.2% Score 258; DB 1; Length 3010;
Best Local Similarity 31.1%; Pred. No.5.5e-09;
Matches 85; Conservative 29; Mismatches 69; Indels 90; Gaps 11;

OY      23 RPAVIPPREVLYQGFDEMEE-AASHLPYTEQGMLAEQFKOKALAF---ASRKNHVSPTHY 78
       :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      1691 KRAVPBPREVLYQGFDEMEBCASNLPTIEQGMQLAEQFKKALLLTQTATGDA----- 1745
OY      79 VPESDARPAIIPREVLRHEFDEMEAPAS-RGNHVSAPHYVPSDSQAAPYLEQ---- 133
Db      1746 -----AAP-----VVESKRRTLEAFAWMNMNFISIGIQYLAGLSTLPGRPAIASMAF 1793
OY      134 -AQVIHQPFEXKALFASRGNHDPPTYHVPSDAKPR-----I 170
Db      1794 TASTSPULTTOSTLLFNILGG-----WVAAGLARPPAASAFLVGAGIAGAIVSIGLVK 1847
OY      171 IPDR-----EVLYREFPEMEESSHL--PYIEQGMVL----- 199
Db      1848 LVDNVAGYGAVGALVALFAFKVMSEMFSTDENVLLRALISPLGLVGVCAILLRRHVD 1907
OY      200 ---AEQFKOKALFASRGNHVADPHTHYTESDA 228
Db      1908 PGEAGVOMNRLIAFASRGNHVSPTHYVESDA 1940

RESULT 3
POLG_HCVRK STANDARD; PRT; 3010 AA.
ID POLG_HCVRK
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepactivin)
DE (EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P55); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus [isolate BK] (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CX NCBI_TaxId=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140696; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by CAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Perge H.E., Wickersham J.A., Hostomsky Z., Hebluka N.,
RA Wooman E.W., Adachi T., Hostomska Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=96227846; PubMed=8568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinhuebler C., Tomei L., de Francesco R., Xuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
```



RL Protein Sci. 7:837-847(1998).  
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +  
CC (RNA)(N).  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: MS8135; AAA72945.1; -  
DR PIR: A38465; GNVVTC.  
DR PDB: 1A10; 25-MAR-98.  
DR PDB: 1JXP; 14-JAN-98.  
DR PDB: 1NS3; 08-APR-98.  
DR PDB: 1CZP; 15-NOV-00.  
DR PDB: 1CSJ; 08-NOV-99.  
DR PDB: 1GX5; 09-APR-02.  
DR PDB: 1GX6; 10-APR-02.  
DR PDB: 1QUV; 26-JUN-00.  
DR PDB: 8OHM; 20-APR-99.  
DR MEROPS: S29.001; -  
DR -----  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002522; HCV\_capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RdRp.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00998; Viral\_RdRp; 1.  
DR Pfam: PD186062; HCV\_NS1; 1.  
DR SMART: SM00487; DEXDC; 1.  
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
KW 3D-structure.  
FT INIT\_MET 1 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
REMOVED FROM CAPSID PROTEIN C BY THE  
CELLULAR AMINOPEPTIDASE.  
CAPSID PROTEIN C (POTENTIAL).  
MATRIX PROTEIN (POTENTIAL).  
MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1616 1662  
FT CHAIN 1663 2013  
FT CHAIN 2014 3010  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1316 1319  
FT CARBOHYD 136 196  
FT CARBOHYD 209 209  
FT CARBOHYD 224 234  
FT CARBOHYD 250 250  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 532 532  
FT CARBOHYD 540 540  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
FT CARBOHYD 2270 2270  
FT CARBOHYD 2240 2240  
FT CARBOHYD 2529 2529  
FT STRAND 2788 2788  
FT STRAND 1031 1035  
FT HELIX 1039 1047  
FT STRAND 1050 1050  
FT STRAND 1059 1063  
FT STRAND 1068 1074  
FT TURN 1075 1076  
FT STRAND 1077 1081  
FT HELIX 1082 1085  
FT TURN 1086 1087  
FT STRAND 1090 1092  
FT TURN 1093 1094  
FT STRAND 1095 1097  
FT STRAND 1101 1103  
FT TURN 1104 1107  
FT STRAND 1108 1112  
FT STRAND 1120 1120  
FT STRAND 1122 1122  
FT STRAND 1129 1133  
FT TURN 1135 1136  
FT STRAND 1139 1144  
FT STRAND 1149 1157  
FT HELIX 1158 1161  
FT TURN 1162 1163  
FT TURN 1165 1166  
FT STRAND 1168 1171  
FT TURN 1172 1174  
FT STRAND 1175 1186  
FT TURN 1187 1188  
FT STRAND 1189 1197  
FT HELIX 1198 1204  
FT TURN 1203 1204  
FT STRAND 1203 1204  
FT STRAND 1680 1688  
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRC64;  
Query Match 14.1%; Score 256; DB 1; Length 3010;  
Best Local Similarity 31.1%; Pred. No. 7,4e-09;  
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;  
QY 23 RPAAIPDREVLVYQFDEMEB-ASHLPYIEQGMQALAEQFKQALAF---ASRQNHVSPTHY 78  
DB 1691 RPAIVPDELLVYQFDEMEB-CASHLPYIEQGMQALAEQFKQALGLIQTATQKAAE----- 1745  
QY 79 VPESDARAALIPDREVLHREFDEMEB-AFASR-GHVPAPAHVPESDASQAPFYEQ----- 133

Db 1746 -----AAP-----VVESKRALFTPMKHMNFISGLOYLAGLTLFGNPAIASLMAF 1793  
Qy 134 -AQVIAHOFKEXLAF-----ASRGNDSPTHVYPPSDAKPAI----- 170  
Db 1794 TASTSPPLTQSTLNLNIGWVAQAQLAPSAASAVGAGIAAAGSIGLKVLYDILA 1853  
Qy 171 -----IPDREVLRYREDEMEESQHL--PYIEGQMML-----AE 201  
Db 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISPGALVGVCAAILRHVGPGECAV 1913  
Qy 202 OFKOKALAFASRGNHVAPTHVYTESDA 228  
Db 1914 QMNRLLAFASRGNHVSPTHVYTESDA 1940  
RESULT 4  
POLG\_HCVUT STANDARD: PRT; 3010 AA.  
AC 000269;  
01-APR-1993 (Rel. 25, Created)  
01-APR-1993 (Rel. 25, Last sequence update)  
15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polypeptide (contains: Capsid protein C (Core protein) (P22);  
Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2  
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepaticivirin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).  
DE Hepatitis C virus (isolate HC-UT) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OC NCBI\_Taxid=31642;  
RN (1)  
SEQUENCE FROM N.A.  
RX MEDLINE:92295714; PubMed:1318627;  
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;  
RT "Molecular cloning of hepatitis C virus genome from a single Japanese  
RT carrier: sequence variation within the same individual and among  
RT infected individuals."  
RT Virus Res. 23:59-53 (1992).  
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polypeptide, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA) (N).  
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA.  
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: D11168; BAA01943.1; -  
DR PIR: A45573; A45573.  
DR PDB: 1A1Q; 2S-MAR-98.  
DR PDB: 1UXP; 14-JAN-98.  
DR MEROPS: S29.001; -  
DR MEROPS: U39.001; -  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002542; HCV\_capsid.  
DR InterPro: IPR002521; HCV\_core.

DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4A.  
DR InterPro: IPR001490; HCV\_NS4B.  
DR InterPro: IPR002868; HCV\_NS5A.  
DR InterPro: IPR002166; HCV\_RDRP.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4A; 1.  
DR Pfam: PF01001; HCV\_NS4B; 1.  
DR Pfam: PF01506; HCV\_NS5A; 1.  
DR Pfam: PF00271; Helicase\_C; 1.  
DR Pfam: PF00998; Viral\_RDRP; 1.  
DR ProDom: PD186062; HCV\_NS1; 1.  
DR SMART: SM00487; DEXDC; 1.  
DR Polypeptide; Glycoprotein; 1.  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;  
KW 3D-structure.  
FT INT\_MET 1 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3010  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1316 1319  
FT CARBOHYD 1316 1319  
FT CARBOHYD 209 209  
FT CARBOHYD 234 234  
FT CARBOHYD 250 250  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 532 532  
FT CARBOHYD 540 540  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
FT CARBOHYD 2077 2077  
FT CARBOHYD 2240 2240  
FT CARBOHYD 2529 2529  
FT CARBOHYD 2788 2788  
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;  
Query Match 14.1%; Score 256; DB 1; Length 3010;  
Best Local Similarity 31.5%; Pred. No. 7,4e-09;  
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;  
Qy 23 RPAAVPPREVLYQFDEMEE-ASHLPYIEGQMQLAEQFKOKALAF---ASRQNHVSPTHY 78  
Db 1691 RPAAVPPREVLYQFDEMEECASHLPYIEGQMQLAEQFKOKALGLLOTATKQ----- 1742

OY 79 VPESDAPPAIIPDREVLRHFEDENEAFASR-GNHVSPAHYVESDASQAAPTEQ---- 133  
 DB 1743 ---AEEAAPVVESE-----WRALFAFWAGKMMNFISGIQYLAGLSTLPQNPALIASLMAF 1793  
 OY 134 -AOVIAHQFEKXVLA-----ASRGNHDSPTHYVESDAPAI----- 170  
 DB 1794 TASITSPLTONTLLFNLGQWVAOLAPPSAASAFVAGAGIAAGAGISGLKVLVDILA 1853  
 OY 171 -----IPDREVLRHFEDENEAFASR-GNHVSPAHYVESDASQAAPTEQ----AE 201  
 DB 1854 GYGAGVAGALVARKWNGSEAPSAEDLVNLLPALSALVAVGVCAILRRHVGPBGAV 1913  
 OY 202 QFQKALAFASRGNHVAPTHYVESDA 228  
 DB 1914 QMNRRLIAFASRGNHVSPTHYVESDA 1940  
 RESULT 5  
 POLG\_HCV1 STANDARD; PRT; 3011 AA.  
 P2664;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polypeptide (Contains: Capsid protein C (core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate 1) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91172826; PubMed=1848704;  
 RA Choo Q.-B., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,  
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,  
 RA Bradley D.W., Kuo G., Houghton M.,  
 RT "genetic organization and diversity of the hepatitis C virus.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC THE SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION  
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -  
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS  
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY  
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL  
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M62321; AAA45676.1; -.  
 DR PIR: A39166; GNMVC3.  
 DR PDB: 1AIV; 16-FEB-99.  
 DR MEROPS: S29.001; -.  
 DR MEROPS: U39.001; -.  
 DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR004109; HCV NS3.  
 DR InterPro: IPR000745; HCV NS4A.  
 DR InterPro: IPR001490; HCV NS4B.  
 DR InterPro: IPR002868; HCV NS5A.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase C.  
 DR InterPro: IPR007095; RNA pol Ds ps.  
 DR InterPro: IPR007094; RNA pol psvit.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4A; 1.  
 DR Pfam: PF01001; HCV NS4B; 1.  
 DR Pfam: PF01506; HCV NS5A; 1.  
 DR Pfam: PF00271; helicase C; 1.  
 DR Pfam: PF00998; viral RdRp; 1.  
 DR Pfam: PF0186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR POLYPROTEIN; Glycoprotein; 1.  
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 DR Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;  
 DR 3D-structure. 1  
 FT INIT\_MET 1  
 FT CHAIN 1  
 FT CHAIN 115  
 FT CHAIN 116  
 FT CHAIN 191  
 FT CHAIN 192  
 FT CHAIN 383  
 FT CHAIN 384  
 FT CHAIN 729  
 FT CHAIN 730  
 FT CHAIN 1006  
 FT CHAIN 1007  
 FT CHAIN 1615  
 FT CHAIN 1616  
 FT CHAIN 1862  
 FT CHAIN 1863  
 FT CHAIN 2014  
 FT CHAIN 3011  
 FT TRANSMEM 347  
 FT ACT\_SITE 1083  
 FT ACT\_SITE 1083  
 FT ACT\_SITE 1107  
 FT ACT\_SITE 1107  
 FT ACT\_SITE 1165  
 FT ACT\_SITE 1165  
 FT NP\_BIND 1230  
 FT NP\_BIND 1237  
 FT SITE 1316  
 FT SITE 1319  
 FT CARBOHYD 196  
 FT CARBOHYD 196  
 FT CARBOHYD 209  
 FT CARBOHYD 209  
 FT CARBOHYD 234  
 FT CARBOHYD 234  
 FT CARBOHYD 305  
 FT CARBOHYD 305  
 FT CARBOHYD 417  
 FT CARBOHYD 417  
 FT CARBOHYD 423  
 FT CARBOHYD 423  
 FT CARBOHYD 430  
 FT CARBOHYD 430  
 FT CARBOHYD 448  
 FT CARBOHYD 448  
 FT CARBOHYD 476  
 FT CARBOHYD 476  
 FT CARBOHYD 532  
 FT CARBOHYD 532  
 FT CARBOHYD 540  
 FT CARBOHYD 540  
 FT CARBOHYD 556  
 FT CARBOHYD 556  
 FT CARBOHYD 576  
 FT CARBOHYD 576  
 FT CARBOHYD 623  
 FT CARBOHYD 623  
 FT CARBOHYD 645  
 FT CARBOHYD 645  
 FT CARBOHYD 2041  
 FT CARBOHYD 2041  
 FT CARBOHYD 2077  
 FT CARBOHYD 2077  
 FT CARBOHYD 2240  
 FT CARBOHYD 2240  
 FT CARBOHYD 2364  
 FT CARBOHYD 2364  
 FT CARBOHYD 2789  
 FT CARBOHYD 2789  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;  
 Query Match 14.0%; Score 254; DB 1; Length 3011;  
 Best Local Similarity 31.8%; Pred. No. 1e-08;  
 Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY 23 RPAVIREVLYOEFDMEAS-HLPYIEQGMOLAEQFKOKALAF---ASRGNHVSPTHY 78  
 Db 1691 KPAIPREVLYREFDEMECSOHLPIEQGMMLAEQFKOKALGLQTA8RQAEV----- 1745  
 QY 79 VPESDARPAIIPREVLYREFDEMEAFASR-GNHVSPPHYVESDASQAAPITEQ----- 133  
 Db 1746 ----TAPV-----QTNWOKLETFWACHMNFISQIYLAGLSTLPNGNPAIASIMAF 1793  
 QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVESDAPAI----- 170  
 Db 1794 TAAVTSPLTTSQTLFPIILGQWVAQAOLAPGAATAFVAGLAGAAGVGLKVLIDILA 1853  
 QY 171 -----IPDREVLYREFDEMECSOHL---PYIEQGMML-----AE 201  
 Db 1854 GYGAGVAGALVARKINSGEVPTEDLVNLLPALSPCALVGVVCAILRRHYVPGGAV 1913  
 QY 202 QFKOKALAFASRGNHVAPTHYTESDA 228  
 Db 1914 QMNRLIAFASRGNHVSPTHYVESDA 1940

RESULT 6  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 ID POLG\_HCVH  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepatitisin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11108;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RA MEDLINE=92052256; Pubmed=1658800;  
 RX Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 [2]  
 X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 h. MEDLINE=97331322; Pubmed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain";  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; Pubmed=9493270;  
 RA Kim J.L., Morgensstern K.A., Grifflith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 RT of unwinding";  
 RL Structure 6:89-100(1998).  
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVITY ROLE.  
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MNNA.  
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M67463; AAA45534.1; -;  
 CC PIR; A36814; GNMVCH.  
 CC PDB; 1HE1; 25-NOV-98.  
 CC PDB; 1AIV; 16-FEB-99.  
 CC PDB; 1AIR; 17-JUN-98.  
 CC MEROPS; S29.001; -;  
 CC MEROPS; U39.001; -;  
 CC TRNSPAC; T04155; -;  
 CC InterPro; IPR001410; DEAD.  
 CC InterPro; IPR002522; HCV\_capsid.  
 CC InterPro; IPR002521; HCV\_core.  
 CC InterPro; IPR002519; HCV\_env.  
 CC InterPro; IPR002531; HCV\_NS1.  
 CC InterPro; IPR002518; HCV\_NS2.  
 CC InterPro; IPR004109; HCV\_NS3.  
 CC InterPro; IPR000745; HCV\_NS4A.  
 CC InterPro; IPR001490; HCV\_NS4B.  
 CC InterPro; IPR002868; HCV\_NS5A.  
 CC InterPro; IPR002166; HCV\_RdRP.  
 CC InterPro; IPR001650; Helicase\_C.  
 CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 CC InterPro; IPR007094; RNA\_pol\_PSVit.  
 CC Pfam; PF01543; HCV\_capsid; 1.  
 CC Pfam; PF01542; HCV\_core; 1.  
 CC Pfam; PF01539; HCV\_env; 1.  
 CC Pfam; PF01560; HCV\_NS1; 1.  
 CC Pfam; PF01538; HCV\_NS2; 1.  
 CC Pfam; PF02907; HCV\_NS3; 1.  
 CC Pfam; PF01006; HCV\_NS4a; 1.  
 CC Pfam; PF01001; HCV\_NS4b; 1.  
 CC Pfam; PF01506; HCV\_NS5a; 1.  
 CC Pfam; PF00271; helicase\_C; 1.  
 CC Pfam; PF00998; Viral\_RdRP; 1.  
 CC ProDom; PD186062; HCV\_NS1; 1.  
 CC SMART; SM00487; DEXDC; 1.  
 CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 CC Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;  
 CC 3D-structure.  
 CC INIT\_MER 1 1  
 CC CHAIN 1 191  
 CC CHAIN 192 383  
 CC CHAIN 384 746  
 CC CHAIN 747 809  
 CC CHAIN 810 1026  
 CC CHAIN 1027 1657  
 CC CHAIN 1658 1711  
 CC CHAIN 1712 1972  
 CC CHAIN 1973 2420  
 CC CHAIN 2421 3011  
 CC TRANSMEM 347 369  
 CC ACT\_SITE 1083 1083  
 CC ACT\_SITE 1107 1107  
 CC ACT\_SITE 1165 1165  
 CC REMOVED FROM CAPSID PROTEIN C BY THE  
 CC CELLULAR AMINOPEPTIDASE.  
 CC CAPSID PROTEIN C.  
 CC ENVELOPE GLYCOPROTEIN E1.  
 CC ENVELOPE GLYCOPROTEIN E2.  
 CC PROTEIN P7.  
 CC NONSTRUCTURAL PROTEIN NS2.  
 CC PROTEASE/HELICASE NS3.  
 CC NONSTRUCTURAL PROTEIN NS4A.  
 CC NONSTRUCTURAL PROTEIN NS4B.  
 CC NONSTRUCTURAL PROTEIN NS5A.  
 CC NONSTRUCTURAL PROTEIN NS5B.  
 CC POTENTIAL.  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).

FT	NP BIND	1230	1237	ATP (POTENTIAL).
FT	SITE	1316	1319	DECH BOX.
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	476	476	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	STRAND	1224	1226	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	TURN	1232	1233	
FT	HELIX	1236	1238	
FT	TURN	1239	1246	
FT	HELIX	1247	1248	
FT	STRAND	1251	1255	
FT	HELIX	1258	1271	
FT	TURN	1272	1272	
FT	STRAND	1277	1280	
FT	TURN	1281	1282	
FT	STRAND	1283	1285	
FT	STRAND	1291	1295	
FT	HELIX	1296	1301	
FT	TURN	1302	1303	
FT	STRAND	1312	1316	
FT	TURN	1317	1319	
FT	HELIX	1323	1335	
FT	TURN	1336	1340	
FT	STRAND	1343	1347	
FT	TURN	1352	1353	
FT	STRAND	1360	1361	
FT	STRAND	1362	1366	
FT	STRAND	1368	1368	
FT	STRAND	1373	1375	
FT	TURN	1376	1377	
FT	STRAND	1378	1380	
FT	HELIX	1382	1385	
FT	STRAND	1389	1393	
FT	HELIX	1397	1409	
FT	TURN	1410	1411	
FT	STRAND	1414	1417	
FT	TURN	1419	1420	
FT	STRAND	1432	1436	
FT	TURN	1438	1439	
FT	STRAND	1450	1453	
FT	STRAND	1456	1463	
FT	STRAND	1471	1478	
FT	STRAND	1480	1480	
FT	HELIX	1481	1488	
FT	TURN	1489	1490	
FT	STRAND	1497	1501	
FT	STRAND	1507	1507	
FT	STRAND	1511	1511	
FT	HELIX	1512	1527	
FT	HELIX	1534	1544	
FT	STRAND	1550	1550	
FT	HELIX	1555	1564	
FT	HELIX	1570	1578	
FT	TURN	1579	1580	
FT	HELIX	1584	1597	
FT	TURN	1598	1598	
FT	HELIX	1606	1611	
FT	TURN	1614	1618	
FT	STRAND	1622	1623	
FT	STRAND	1627	1627	
FT	STRAND	1635	1636	

```

FT      HELIX      1640 1652
SO      SEQUENCE   3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match      13.6%; Score 247; DB 1; Length 3011;
Best Local Similarity 32.2%; Pred. No. 2; 8e-08;
Matches 86; Conservative 26; Mismatches 77; Indels 78; Gaps 11;

QY      23 RPAVIDPDEVLYQFDEMEBAS-HLPYIEGQMLAEQFKOKALAF---ASRONHVSPTHY 78
        1691 KPAIIPDEVLYQFDEMECSOHLPIEQGMMLAEQFKOKALGLQTAASHAEV----- 1745
        79 VPESDAPRAIIPDEVLYHREFDEMEEPASR-GNHVSPAHVYESDASQAPYIEQ---- 133
        1746 -----ITPAV-----QTNMQLKLEFWAGHMNFISGIQYLAGLSTLPGNPAAIASIMAF 1793
QY      134 -AQVIAHOFKEKYLAF-----ASRGNDSPTHVVPESDAKPAI----- 170
        1794 TAAVTSPLTTGQTLLENILGWAALAPGAATAFVAGLACALADSVGLKAVLYDILA 1853
QY      171 -----IPDEVLYREFDEMEESQH-LPYI-----EQGMMLAE----- 201
        1854 GYGAGVAGALVAFKINSGEVPTEDLVNLLPALISGALAVGVVFAISLRRYGPEGAV 1913
QY      202 QFKOKALPASPGRGNHVAPTHYVESDA 228
        1914 QNMNRLIAFPASRGNHVSPHYVESDA 1940

Db      1914 QNMNRLIAFPASRGNHVSPHYVESDA 1940

RESULT 7
POLG_HCVJ6
ID      POLG_HCVJ6      STANDARD;      PRT; 3033 AA.
AC      P26660;
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Genome polyprotein (containing: Capsid protein C (Core protein) (P22);
DE      Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
DE      (GP68) (GP70) (NS1); Protease/helicase NS3 (P70) (Hepatitisin)
DE      (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE      (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE      NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE      NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
OS      Hepatitis C virus (isolate HC-06) (HCV).
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_Taxid=11113;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9204440; PubMed=1658196;
RA      Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Litzuka H.,
RA      Machida A., Miyakawa Y., Mayumi M.;
RT      "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT      from a human carrier: comparison with reported isolates for conserved
RT      and divergent regions."
RL      J. Gen. Virol. 72:2697-2704(1991).
CC      -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC      HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC      NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC      -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC      precursor polyprotein, commonly with Asp or Glu in the P6
CC      position, Cys or Thr in P1 and Ser or Ala in P1'.
CC      -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC      {RNA} (N).
CC      -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND RNA.
CC      -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way

```

```

FT CARBOHYD 2811 2811 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
Query Match 11.2%; Score 203.5; DB 1; Length 3033;
Best Local Similarity 44.1%; Pred. No. 2e-05;
Matches 60; Conservative 10; Mismatches 37; Indels 29; Gaps 5;

Oy 229 KPALVPDKVELVYQYDEMEENAFASRGNHVAPTHVVESDAS---ASLPYDETRAIAGOF 285
Db 1619 KPTLVGPTPLLYR-----LGSVTNEVTLTHPTVTKYATCNQADLEWNTSTVLAQGV 1670
Oy 286 KEKVLAPASRGN-----HVSFRHVVPESEPOVYVTPDKELLYEAFDEMEF-ASQAL 336
Db 1671 LAAYAAACATGCGVCIIIGRLHVNR-----AVAPDKVELVYEAFADEMEECASRAAL 1721
Oy 337 IEEGORAEMLKSKTIQ 352
Db 1722 IEEGORAEMLKSKTIQ 1737

RESULT 8
POLG_HCVJ8
ID_POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.46)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxId=11115;
RN [1]
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kuzai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 186:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10988; BAA01761.1; .
DR PIR; A40250; GNMVJ8.
DR HSSP; P27958; 1HE1.
DR MEROPS; S29.001; -.

```

MEMOPS: U39.001. -  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5b.  
 DR InterPro: IPR007095; RNA\_pol\_PS.PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 191 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 299 299  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2359 2359  
 FT CARBOHYD 2811 2811  
 SO SEQUENCE 3033 AA; 330177 MW; 1A173E7E381FDDA CRC64;  
 Query Match 11.1%; Score 201.5; DB 1; Length 3033;  
 Best Local Similarity 46.2%; Pred. No. 2.6e-05;  
 Matches 60; Conservative 13; Mismatches 40; Indels 17; Gaps 5;

Db 1619 KPTLTGPPPLR-----LCAVTNEVTLTHPTKTIATCMQADLITSSWLACV 1670  
 QY 286 KEKVLAFASRGNHVS--PRHYVESEPOVVTTPDXEILYEAFDEME-ASKAALIEGQR 342  
 Db 1671 LAVAAYCLATGCTIIRLHND---RVVAPDEKILYEAFDEMEECASKALIEGQR 1727  
 QY 343 MAEMLKSKIQ 352  
 Db 1728 MAEMLKSKIQ 1737  
 RESULT 9  
 ID AAC4 RAT STANDARD; PRT; 911 AA.  
 AC Q9OXQ0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alpha-actinin 4 (Non-muscle alpha-actinin 4) (F-actin cross linking protein).  
 GN ACTN4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=101116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20139748; PubMed=10673389;  
 RA El-Husseini A.E.-D., Kwasnicka D., Yamada T., Hirahashi S.,  
 RA Vincent S.R.;  
 RT "BERP, a novel ring finger protein, binds to alpha-actinin-4";  
 RL Biochem. Biophys. Res. Commun. 267:906-911(2000).  
 CC -!- FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING PROTEIN.  
 CC -!- SUBUNIT: HOMODIMER, ANTIPARALLEL (BY SIMILARITY). BINDS BERP AT THE N-TERMINUS.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Contains 1 actin-binding domain.  
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
 CC -!- SIMILARITY: Contains 4 spectrin repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).  
 CC -----  
 DR EMBL: AF190909; AAF20064.1; -;  
 DR PIR: JC7186; JC7186.  
 DR HSSP: Q01082; 1BKR.  
 DR InterPro: IPR001589; Acbind actin.  
 DR InterPro: IPR001715; Calponin-like.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00307; CH; 2.  
 DR Pfam: PF00036; ethand; 2.  
 DR Pfam: PF00435; spectrin; 4.  
 DR ProDom: PD000012; EF-hand; 1.  
 DR SMART: SM00033; CH; 2.  
 DR SMART: SM00054; EFh; 2.  
 DR SMART: SM00150; SPECT; 3.  
 DR PROSITE: PS00019; ACTININ\_1; 1.  
 DR PROSITE: PS00020; ACTININ\_2; 1.  
 DR PROSITE: PS00021; CH; 2.  
 DR PROSITE: PS00018; EF\_HAND; 1.  
 KW Actin-binding; Calcium-binding; Repeat; Multigene family;  
 FT DOMAIN 1 269 ACTIN-BINDING (BY SIMILARITY).







```

FT REPEAT 1593 1660 SPECTRIN 8.
FT REPEAT 1817 1886 SPECTRIN 9.
FT REPEAT 1934 2044 SPECTRIN 10.
FT REPEAT 2262 2282 SPECTRIN 11.
FT REPEAT 2376 2397 SPECTRIN 12.
FT REPEAT 2400 2509 SPECTRIN 13.
FT REPEAT 2557 2617 SPECTRIN 14.
FT REPEAT 2620 2727 SPECTRIN 15.
FT REPEAT 2730 2837 SPECTRIN 16.
FT REPEAT 2840 2944 SPECTRIN 17.
FT REPEAT 2946 2972 SPECTRIN 18.
FT REPEAT 3029 3029 SPECTRIN 19.
FT REPEAT 3086 3162 SPECTRIN 20.
FT REPEAT 3195 3273 SPECTRIN 21.
FT REPEAT 3276 3382 SPECTRIN 22.
FT REPEAT 3385 3491 SPECTRIN 23.
FT REPEAT 3494 3518 SPECTRIN 24.
FT REPEAT 3603 3709 SPECTRIN 25.
FT REPEAT 3722 3817 SPECTRIN 26.
FT REPEAT 3869 3930 SPECTRIN 27.
FT REPEAT 3933 4039 SPECTRIN 28.
FT REPEAT 4042 4149 SPECTRIN 29.
FT REPEAT 4152 4258 SPECTRIN 30.
FT REPEAT 4261 4368 SPECTRIN 31.
FT REPEAT 4371 4477 SPECTRIN 32.
FT REPEAT 4480 4587 SPECTRIN 33.
FT REPEAT 4594 4695 SPECTRIN 34.
FT REPEAT 4698 4804 SPECTRIN 35.
FT REPEAT 4807 4872 SPECTRIN 36.
FT REPEAT 4917 4941 SPECTRIN 37.
FT CA_BIND 4987 4998 EF-HAND 1 (POTENTIAL).
FT CA_BIND 5023 5034 EF-HAND 2 (POTENTIAL).
FT DOMAIN 5173 5180 4 X 4 AA TANDEN REPEATS OF [GS]-S-R-[AR].
FT VARSPIC 5252 5267 POLY-SER.
MSSDETLSEKRSRSGRSYRSRSGSLSPCPD
TLPWNLPHKOKRKSDSVLDPAEAVAVVADERVQK
LPREKGRMFRHLONVIALDFKOROVKLVNRNDIDG
NPKTLGLIWTITLHFQ -> MENSAGCYKPEKSTAVBEK
APISPKRVAFKAKMKKTLITPASHREALEGTIVEST
ETLTKLARLPKEPGVAGAPPSDIFLPDGSAPNSGVDO
GMIVOVKESFOAEIQTALHLENSSVVGAMDSLEGMTV
IAHLDPARNCEKSVSOLPEPTPASCSSRAVLPLQGE
TAVEOGTLRLHRHRSSTLPRTDYPSETVDODQPSRGVSG
GRKTSVSAPTGSMWIAKCSVASIFKOSGDPHTPTVAG
LVCKGPIMPASOSDLSVSGITVSLPSSSGVSGDRLRHG
LRPDETEPKTSTPFSFEDDTLSLE (in isoform 3).
/FTid=VSP.000718.
L -> P (IN REF. 2; AAC52990).
O -> H (IN REF. 2)
Query Match 6.6%; Score 120.5; DB 1; Length 5327;
Best Local Similarity 19.0%; Pred. No. 9.9;
Matches 80; Conservative 76; Mismatches 124; Indels 141; Gaps 19;

```

```

Db 3696 VLLNSERYARLERAOVLNQFWE-----TYEELSPMAETLALIAQLPPAVDHE 3746
Qy 176 VLREDEME-----SCHLPYIEOGMLAEQFKOKALAFASRGNHVAPTHVTESDAKP 230
Db 3747 QLRQOEERQRLRESIAEHKPHIDIKIKIPOLKE----- 3781
Qy 231 ALVPDK-EVLYOQYDEMEAEFASRGNHVAPTHVVESDASASLPYMETRAIAGQFKEXV 289
Db 3782 -LNPEKGKVEEKYOKAENMYAQ-----IKDEVRORALDEAVSQSQAFHDKI 3829
Qy 290 LAFASRGNHVSPRHVPESEPGVVTPPKELIYEFDEMEASKALIEGGRMAEMLS 349
Db 3830 EPMLETLNLSRLMP---PLIPAEVK--IRECISDNKXYT--VELEKLPSEFALKR 3882
Qy 350 K 350
Db 3883 R 3883

RESULT 11
AAC4 MOUSE STANDARD: PRT: 912 AA.
ID AAC4 MOUSE
AC P57780:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-actinin 4 (Non-muscle alpha-actinin 4) (f-actin cross linking
protein).
GN ACTN4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=20422668; PubMed=10964513;
RA Dear T.N., Meier N.T., Hunn M., Boehm T.;
RT "Gene structure, chromosomal localization and expression pattern of
RL Capn12, a new member of the calpain large subunit gene family.";
Genomics 69:152-160(2000).
CC - FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR
CC ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING
CC PROTEIN.
CC - SUBUNIT: HOMODIMER, ANTIPARALLEL (BY SIMILARITY).
CC - SIMILARITY: Contains 1 actin-binding domain.
CC - SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC - SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC - SIMILARITY: Contains 4 spectrin repeats.
CC - SIMILARITY: Contains 4 spectrin repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
CC EMBL: AJ269242; CAC10069.1; -
CC DR HSSP; 001082; 1BKR.
CC DR MGD; MGI:1890773; Actn4.
CC DR InterPro; IPR001589; Actbind actuin.
CC DR InterPro; IPR001715; Calponin-like.
CC DR InterPro; IPR002048; EF-hand.
CC DR InterPro; IPR002017; Spectrin.
CC DR Pfam; PF00307; CH; 2.
CC DR Pfam; PF00036; ehand; 2.
CC DR Pfam; PF00435; spectrin; 4.
CC DR ProDom; PD000012; EF-hand; 1.
CC DR SMART; SM00054; CH; 2.
CC DR SMART; SM00054; EFH; 2.
CC DR SMART; SM00150; SPC; 3.

```

```

DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS00021; CH; 2.
KW Actin-binding; Calcium-binding; Repeat; Multigene family.
FT DOMAIN 1 267 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 1 155 CH 1.
FT DOMAIN 164 267 CH 2.
FT DOMAIN 178 193 POLYPHOSPHOINOSITIDE (PIP2)-BINDING
(POTENTIAL).
FT REPEAT 294 404 SPECTRIN 1.
FT REPEAT 414 519 SPECTRIN 2.
FT REPEAT 529 640 SPECTRIN 3.
FT REPEAT 650 753 SPECTRIN 4.
FT CA_BIND 779 790 EF_HAND 1 (POTENTIAL).
FT CA_BIND 820 831 EF_HAND 2 (POTENTIAL).
SQ SEQUENCE 912 AA; 104976 MW; 07AA9C92AC228B5A CRC64;

Query Match 6.3%; Score 114.5; DB 1; Length 912;
at Local Similarity 19.9%; Pred. No. 2.8;
tches 80; Conservative 65; Mismatches 124; Indels 133; Gaps 20;

QY 7 LEQGHLEAFOFKOKA-----LRPAVIPDREVLVYQEFDEMEE--ASHL 46
DB 409 LERLDHLAEKFRKOKASIHAWTDGKEAMLKORDYETATLSDIKALIRKHEAFESDLAAHQ 468
QY 47 PYIEQOMLAEOFKO-----KALAFASRGNHV--SPHHYVESDARPAIIPDREV 94
DB 469 DVEEQALAIQENEDLDYDSHNNTROCKICDQMDNLSLTH-----SRKEALEKTEK 522
QY 95 LHRFEDMEAFASRGNHVSPAHYVESDASQAAPYIEQAOVIAHOFKEKVLAFASRGNH 154
DB 523 QLETTDQHLLEYAKRA-----APFNMMESAMEDLDQMFVHTI -EIGGLISA--H 571
QY 155 DSPTHVPPSD-AKPAIIPDREVLVREFEMESQHL-----PY-----I 193
DB 572 DQFKSTLPADDEREMIL-----AIHKEAQRIAESNHIKLSGSPYTTVTPQIINSKWEKV 627
QY 194 EQ-----GMMLAEQFKOKA-----LAFASRGNHVAP----- 219
DB 628 QQLVPRKDHALLEBQSKQSNHLEHRQFASQANWGPWIOTKMEIIGRISIMNGTLEQD 687
QY 220 -THVYTES---DAKPALVPDKEVLVYQYDEMEEAFAASRGNHVAPTHVVE----- 265
DB 688 LSHLQYERSIVDYKPSL---DLLEQHQQLIQEALIFPNKH---TNYIMEHIRGVMEQL 740
QY 266 -SDASASLPYME---TRAIACQFKEKVLAFASRGNHVSPRH 303
DB 741 LTTIARTINEVENQILTRDAKGISQEQMGEFRASFHFDKDH 782

RESULT 12
PLE1_CRIGR STANDARD; PRT; 4473 AA.
AC 09J15;
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Plectin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated
protein) (IFAP300) (Fragment).
GN PLECT1.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20334248; PubMed=10873583;
RA Clubb B.H., Chou Y.-H., Herrmann H., Svickina T.M., Borisy G.G.,
RA Goldman R.D.;
RT "The 300-kDa intermediate filament-associated protein (IFAP300) is a
hamster plectin ortholog.";

```

```

RL Biochem. Biophys. Res. Commun. 273:183-187(2000).
RN (2)
RX PHOSPHORYLATION.
RX MEDLINE=96215219; PubMed=8626512;
RA Malecz N., Foisner R., Stadler C., Wiche G.;
RT "Identification of plectin as a substrate of p34cdc2 kinase and
RT mapping of a single phosphorylation site.";
RL J. Biol. Chem. 271:8203-8208(1996).
CC -!- FUNCTION: INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESOMES OR
HEMIDESOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN. THE C-TERMINUS WITH
VIMENTIN, DESMIN, GAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
AND THE C-TERMINUS CAN BIND INTERFERON BETA-4.
CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
INTERMEDIATE FILAMENTS DURING MITOSIS.
CC -!- SIMILARITY: Contains 1 actin-binding domain.
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -!- SIMILARITY: Contains 32 plectin repeats.
CC -!- SIMILARITY: Contains 4 spectrin repeats.
CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF260753; AAF70372.1; -.
DR HSSP; Q01082; 1BRK.
DR InterPro; IPR001589; Actbind_actin.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR001101; Plectin repeat.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00681; Plectin; 20.
DR SMART; SM00033; CH; 2.
DR SMART; SM00250; PLEC; 34.
DR PROSITE; PS00019; ACTININ_1; PARTIAL.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS00021; CH; 2.
KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW Phosphorylation.
FT DOMAIN 1 1 1259 GLOBULAR 1.
FT NON_TER 1 1 CENTRAL FIBROUS ROD DOMAIN.
FT DOMAIN 1260 2544 GLOBULAR 2.
FT DOMAIN 2545 4473 ACTIN-BINDING.
FT DOMAIN <1 192 CH 1.
FT DOMAIN <1 74 CH 2.
FT REPEAT 87 189 SPECTRIN 1.
FT REPEAT 449 508 SPECTRIN 2.
FT REPEAT 529 613 SPECTRIN 3.
FT REPEAT 626 719 SPECTRIN 4.
FT REPEAT 1104 1204 SPECTRIN 5.
FT REPEAT 1258 2548 COILED COIL (POTENTIAL).
FT REPEAT 2615 2652 PLECTIN 1.
FT REPEAT 2653 2690 PLECTIN 2.
FT REPEAT 2691 2728 PLECTIN 3.
FT REPEAT 2729 2766 PLECTIN 4.
FT REPEAT 2767 2804 PLECTIN 5.
FT REPEAT 2805 2942 PLECTIN 6.
FT REPEAT 2943 2980 PLECTIN 7.
FT REPEAT 2981 3018 PLECTIN 8.
FT REPEAT 3019 3056 PLECTIN 9.
FT REPEAT 3057 3094 PLECTIN 10.
FT REPEAT 3095 3274 PLECTIN 11.
FT REPEAT 3275 3312 PLECTIN 12.
FT REPEAT 3313 3387 PLECTIN 13.

```

FT REPEAT 3388 3425 PLECTIN 14.  
 FT REPEAT 3429 3463 PLECTIN 15.  
 FT REPEAT 3609 3646 PLECTIN 16.  
 FT REPEAT 3647 3684 PLECTIN 17.  
 FT REPEAT 3685 3722 PLECTIN 18.  
 FT REPEAT 3723 3760 PLECTIN 19.  
 FT REPEAT 3764 3797 PLECTIN 20.  
 FT REPEAT 3800 3834 PLECTIN 21.  
 FT REPEAT 3852 3889 PLECTIN 22.  
 FT REPEAT 3890 3927 PLECTIN 23.  
 FT REPEAT 3928 3965 PLECTIN 24.  
 FT REPEAT 3966 4003 PLECTIN 25.  
 FT REPEAT 4007 4041 PLECTIN 26.  
 FT REPEAT 4043 4094 PLECTIN 27.  
 FT REPEAT 4197 4234 PLECTIN 28.  
 FT REPEAT 4235 4272 PLECTIN 29.  
 FT REPEAT 4273 4310 PLECTIN 30.  
 FT REPEAT 4311 4348 PLECTIN 31.  
 FT REPEAT 4349 4386 PLECTIN 32.  
 FT REPEAT 4039 4089 BINDING TO INTERMEDIATE FILAMENTS (BY SIMILARITY).  
 FT DOMAIN 4414 4429 4 X 4 AA TANDEN REPEATS OF G-S-R-X.  
 FT MOD RES 4328 4328 PHOSPHORYLATION (BY CDC2).  
 SQ SEQUENCE 4473 AA; 509015 MW; E14615D361E3484 CRC64;  
 Query Match 6.3%; Score 114.5; DB 1; Length 4473;  
 Best Local Similarity 19.6%; Pred. No. 20;  
 Matches 73; Conservative 66; Mismatches 172; Indels 61; Gaps 10;  
 QY 8 ECGHMLAEQFKOKAL-----RPAVIDREVL-----YQFDEMEASHLPYI 49  
 DB 1906 EORREAEERVOXSLAEEDAROKALAEVEERLAKKVEBARRLRERAEQESRQLOLA 1965  
 QY 50 ECGQVLAEOFKOKALAFASRGNHVSPTHYVPESDAPRAIIPDEVLHREFDEMEAFASR 109  
 DB 1966 QEAQAQKRLQAEKNAHVQOR-----EEELQOTLQOEGSMLERLGAEVAA-RRA 2015  
 QY 110 GNVSPHYVPESDASQAPYIEQAVIAH-----QFKKVLAFASRGHSDSPHYVPE 163  
 DB 2016 AEEAEERAEQAEEREAQSRQVFEAEERLKQSAEQAOARQAQAAAEKLEKAEQEMARR 2075  
 QY 164 SDAKPAIIPDEVLHREFDEMEASHLPYIECGHMLAEQFKOKALAFASRGNHVAIPHVV 223  
 DB 2076 AQBEOALKQKQAA-----DAEMEKHKKFAEQTLROKAQVQOELTTLRLQ-----L 2121  
 QY 224 TESDAKPAIIPDEVLHREFDEMEASHLPYIECGHMLAEQFKOKALAFASRGNHVAIPHVV 223  
 DB 2122 EETDQKSL--DEELQRLKAEVTEARQSRQ-----VEELTFVVRQVMEELGKTLKA 2171  
 DB 284 QFKKVLAFASRGNHVSPTHYVPESDAPRAIIPDEVLHREFDEMEAFASR 340  
 DB 2172 RIEAENRALLLRDKNTQRLFEAEERKQVAEERLVAQAQEAARLRQLAEEDLAQOR 2231  
 QY 341 QRMAMLKSKIO 352  
 DB 2232 ALAEKMLKEMQ 2243  
 RESULT 13  
 AAC4\_HUMAN STANDARD; PRT; 911 AA.  
 ID AAC4\_HUMAN  
 AC Q43707; 076048;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alpha-actinin 4 (Non-muscle alpha-actinin 4) (F-actin cross linking protein).  
 GN ACTN4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE OF 4-911 FROM N.A.  
 RX MEDLINE=9817716; PubMed=9508771;  
 RA Honda K., Yamada T., Endo R., Ino Y., Gotoh M., Tsuda H., Yamada Y.,  
 RA Chiba H., Hirohashi S.;  
 RT "Actinin-4, a novel actin-bundling protein associated with cell  
 RT motility and cancer invasion.";  
 RL J. Cell Biol. 140:1383-1393(1998).  
 RN [2]  
 RP ERRATUM.  
 RA Honda K., Yamada T., Endo R., Ino Y., Gotoh M., Tsuda H., Yamada Y.,  
 RA Chiba H., Hirohashi S.;  
 RL J. Cell Biol. 143:276-276(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroblastoma;  
 RX MEDLINE=20120478; PubMed=10656685;  
 RA Nikolopoulos S.N., Spengler B.A., Kisselbach K., Evans A.E.,  
 RA Biedler J.L., Ross R.A.;  
 RT "The human non-muscle alpha-actinin protein encoded by the ACTN4 gene  
 RT suppresses tumorigenicity of human neuroblastoma cells.";  
 RL Oncogene 19:380-386(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustun T.B., Toshlyuk S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,  
 RA Bhasak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs S.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Maira M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP SEQUENCE OF 1-170 FROM N.A.  
 RA Ottensmeyer F., Obermaier B., Mewes H.-W., Gassenhuber J.,  
 RA Wiemann S.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 1-218 FROM N.A.  
 RA Isogai T., Otsuki T., Sugiyama T.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP VARIANTS FSGS1 GIU-255; ILB-259 AND PRO-262.  
 RC TISSUE=Lymphocytes;  
 RX MEDLINE=20164321; PubMed=10700177;  
 RA Kaplan J.M., Kim S.H., North K.N., Renke H., Correia L.A.,  
 RA Tong H.-Q., Mathis B.J., Rodriguez-Perez J.-C., Allen P.G.,  
 RA Beggs A.H., Pollak M.R.;  
 RT "Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal  
 RT segmental glomerulosclerosis.";  
 RL Nat. Genet. 24:251-256(2000).  
 CC -!- FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR  
 CC ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING  
 CC PROTEIN.  
 CC -!- SUBUNIT: HOMODIMER, ANTIPARALLEL (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOSOLIC. COLOCALIZES WITH  
 CC ACTIN STRESS FIBERS. NUCLEAR TRANSLLOCATION CAN BE INDUCED BY THE  
 CC PI3 KINASE INHIBITOR MORTALININ OR BY CYTOCHALASIN D. EXCLUSIVELY  
 CC LOCALIZED IN THE NUCLEUS IN A LIMITED NUMBER OF CELL LINES (BREAST  
 CC CANCER CELL LINE MCF7, ORAL FLOOR CANCER IMC2, AND BLADDER CANCER

Query Match	Similarity	Score	DB 1	Length	911
Best Local	79	6.24	Conservative	66	Mismatches 132; Indels 117; Gaps 19;
FT	REPEAT	649	752	SPECTRIN 4.	
FT	CA_BIND	778	789	EF-HAND 1 (POTENTIAL).	
FT	CA_BIND	819	830	EF-HAND 2 (POTENTIAL).	
FT	DOMAIN	19	26	POLY-GLY.	
FT	VARIANT	255	255	K -> E (in FSGS1).	
FT	VARIANT	259	259	/FtId=VAR_010378.	
FT	VARIANT	259	259	T -> I (in FSGS1).	
FT	VARIANT	262	262	/FtId=VAR_010379.	
FT	CONFLICT	60	60	S -> P (in FSGS1).	
FT	CONFLICT	124	124	C -> S (in REF. 3).	
FT	CONFLICT	164	164	V -> I (in REF. 3).	
FT	CONFLICT	276	276	S -> L (in REF. 4).	
FT	CONFLICT	292	294	T -> TET (in REF. 3).	
FT	CONFLICT	359	360	EHL -> CSTS (in REF. 3).	
FT	CONFLICT	476	476	TL -> SV (in REF. 3).	
FT	CONFLICT	526	526	I -> S (in REF. 3).	
FT	CONFLICT	536	536	I -> I (in REF. 3).	
FT	CONFLICT	645	645	R -> P (in REF. 3).	
FT	CONFLICT	673	674	Q -> QQ (in REF. 3).	
FT	CONFLICT	850	850	GR -> A (in REF. 3).	
FT	CONFLICT	859	859	A -> T (in REF. 3).	
FT	CONFLICT	891	893	MISSING (in REF. 3).	
FT	CONFLICT	891	893	AVP -> GVR (in REF. 3).	
FT	SEQUENCE	911 AA	104854 MW	461580C3F22937D1 CRC64;	
Db	7	LEQGMLAEQFKOKA	-----	LRPAVDPREVLYQEFDEMEE	ASHL 46
Db	408	LERLDHLAEKFRQKASIEHAWTDQKEAMLKRDYETATLSLDIKALIRKHEAFESDLAAHQ	467		
Qy	47	PYIEQSGQGLAEQFKOKALAFASQNHVSPTHYVPE	DARPAIIPD	REVLYHR	-----E 98
Db	468	DREYQALIAQELNE	LDYDSDHNVTTRCQKICDQDAAGSLTSHSRREALETEKOLEA	525	
Qy	99	FDEMEAFASRGNHVSPAHVVPESDSQAAPRYEQAVLHQREKYLAFASRGNHDSPT	158		
Db	526	IDQLHLEYAKRA	-----	APRNMMESAMEDLOMFYHTI	EIEGLISA--HDQK 574
Qy	159	HYVPESD-AKPAIIPDREVLYREFDEMEEQHL	-----PY	-----	IEQ-- 195
Db	575	STLPDADREREAL	-----	AIHKEAQRIAEENHIIKLSQSNYTTVTPQIINSKMEKQQLV	630
Qy	196	-----GMLLAQFOFKQA	-----	LAFASRGNHVAP	-----
Db	631	PKEDHALLBEOSSKQOSNEHLRRFOFASQANVVGWIGIOTKMEIGRISIEMGSTLEDOLSHL	690		
Qy	223	VTESDAKRALVPRPEKVLVYQOYDMEEFASRGNHVAPPTHVVE	-----	SDASASL	272
Db	691	KQYERSLVDPKPVLDLEQOHQIQLDFDNK	-----	TNYTMEHIVGVEQLLTARTI	747
Qy	273	PYME	-----	TRAIAQFKEKYLAFASRGNHVSPPH	303
Db	748	NEVENQILTRDAKGISQEQWQEFASRNHHDKDH	781		

RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Placenta;  
 RX MEDLINE=96210632; PubMed=8633055;  
 RA Liu C.-G., Maetcker C., Castanon M.J., Hauptmann R., Wiche G.;  
 RT "Human plectin: organization of the gene, sequence analysis, and  
 chromosome localization (8q24)";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.  
 RX MEDLINE=96312447; PubMed=8698233;  
 RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,  
 RA Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owaribe K.,  
 RA McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,  
 RA Utto J.;  
 RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:  
 RT cDNA cloning and genomic organization";  
 RL Genes Dev. 10:1724-1735(1996).  
 RN [3]  
 RP VARIANT MD-EBS 1003-GLN-ALA-1005 DEL.  
 RX MEDLINE=97049959; PubMed=8894687;  
 RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,  
 RA Hachisuka H., Nishikawa T., McLean W.H.I., Utto J.;  
 RT "Homozygous deletion mutations in the plectin gene (PLEC1) in patients  
 RT with epidermolysis bullosa simplex associated with late-onset  
 RT muscular dystrophy";  
 RL Hum. Mol. Genet. 5:1539-1546(1996).  
 RN [4]  
 RP VARIANT MD-EBS LEU-429 INS.  
 RX MEDLINE=21090821; PubMed=1159198;  
 RA Bauer J.W., Rouan F., Kofler B., Reznicek G.A., Kornacker I.,  
 RA Muss W., Hanebner R., Klausegger A., Huber A., Fohla-Gubio G.,  
 RA Wiche G., Utto J., Hutter H.;  
 RT "A compound heterozygous one amino-acid insertion/nonsense mutation in  
 RT the plectin gene causes epidermolysis bullosa simplex with plectin  
 RT deficiency";  
 RL Am. J. Pathol. 158:617-625(2001).  
 RN [5]  
 RP FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND  
 RP MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR  
 RP HEMIDESMOSOMES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO  
 RP MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE  
 RP CROSSLINKING AND STABILIZATION OF CYTOSKELETAL INTERMEDIATE  
 RP FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.  
 RP SUBUNIT: HOMODIMER OR HOMOTETRAMER.  
 RP ALTERNATIVE PRODUCTS:  
 RP Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=Q15149-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q15149-2; Sequence=VSP\_005030;  
 CC Name=3;  
 CC IsoId=Q15149-3; Sequence=VSP\_005030, VSP\_005031;  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN  
 CC MUSCLE, HEART, PLACENTA AND SPINAL CORD.  
 CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH  
 CC VIMENTIN, DESMIN, GAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-  
 CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.  
 CC -1- PTM: PHOSPHORYLATED BY CDC2, REGULATES DISSOCIATION FROM  
 CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
 CC -1- DISEASE: DEFECTS IN PLECT1 ARE THE CAUSE OF EPIDERMOLYSIS BULLOSA  
 CC WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-MD), AN AUTOSOMAL RECESSIVE  
 CC DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL  
 CC OF THE HEMIDESMOSOME AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.  
 CC -1- SIMILARITY: Contains 1 actin-binding domain.  
 CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC -1- SIMILARITY: Contains 3 plectin repeats.  
 CC -1- SIMILARITY: Belongs to the plakin or cyclinker family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z5367; CAA91196.1; -  
 CC EMBL; U53204; AAB05427.1; -  
 CC EMBL; U63610; AAB05428.1; -  
 CC EMBL; U63609; AAB05428.1; JOINED.  
 CC EMBL; X97053; CAA65765.1; -  
 CC PIR; C59404; A59404.  
 CC HSSP; Q01082; 1BRK.  
 CC Genew; HGNC:9069; PLECT1.  
 CC GK; Q15149; -  
 CC MIM; 601282; -  
 CC MIM; 22670; -  
 CC GO; GO:0008307; F:structural constituent of muscle; TAS.  
 CC InterPro; IPR001589; Actbind actin.  
 CC InterPro; IPR001715; Calponin-like.  
 CC InterPro; IPR001101; Plectin\_repeat.  
 CC InterPro; IPR005326; S10\_plectin.  
 CC InterPro; IPR002017; Spectrin.  
 CC Pfam; PF00307; CH; 2.  
 CC Pfam; PF00681; Plectin; 19.  
 CC Pfam; PF03501; S10\_plectin; 1.  
 CC ProDom; PD006662; S10\_plectin\_N; 1.  
 CC SMART; SM00033; CH; 2.  
 CC SMART; SM00250; PLEC; 34.  
 CC PROSITE; PS00019; ACTININ\_1; FALSE\_NEG.  
 CC PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
 CC PROSITE; PS50021; CH; 2.  
 CC Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;  
 CC Phosphorylation; Alternative splicing; Epidermolysis bullosa;  
 CC Disease mutation.  
 KW KW  
 FT DOMAIN 1 1470  
 FT DOMAIN 1471 2755  
 FT DOMAIN 2756 4684  
 FT DOMAIN 175 400  
 FT DOMAIN 179 282  
 FT DOMAIN 295 397  
 FT REPEAT 645 710  
 FT REPEAT 740 824  
 FT REPEAT 837 930  
 FT REPEAT 1315 1415  
 FT DOMAIN 1469 2756  
 FT REPEAT 2826 2863  
 FT REPEAT 2864 2901  
 FT REPEAT 2902 2939  
 FT REPEAT 2940 2977  
 FT REPEAT 2981 3015  
 FT REPEAT 3116 3153  
 FT REPEAT 3154 3191  
 FT REPEAT 3192 3229  
 FT REPEAT 3230 3267  
 FT REPEAT 3268 3305  
 FT REPEAT 3306 3343  
 FT REPEAT 3485 3522  
 FT REPEAT 3523 3560  
 FT REPEAT 3561 3598  
 FT REPEAT 3599 3636  
 FT REPEAT 3640 3674  
 FT REPEAT 3674 3857  
 FT REPEAT 3858 3895  
 FT REPEAT 3896 3933  
 FT REPEAT 3934 3971  
 FT REPEAT 3975 4008  
 FT REPEAT 4063 4100  
 FT REPEAT 4101 4138  
 FT REPEAT 4139 4176  
 FT REPEAT 4177 4214  
 FT REPEAT 4218 4252  
 FT REPEAT 4265 4305  
 FT REPEAT 4319 4356  
 FT REPEAT 4408 4445  
 FT REPEAT 4446 4483  
 FT PLECTIN 30.  
 FT PLECTIN 1. CENTRAL FIBROUS ROD DOMAIN.  
 FT PLECTIN 2. GLOBULAR 2.  
 FT PLECTIN 1. CH 1.  
 FT PLECTIN 1. CH 2.  
 FT SPECTRIN 1.  
 FT SPECTRIN 2.  
 FT SPECTRIN 3.  
 FT SPECTRIN 4.  
 FT COILED COIL (POTENTIAL).  
 FT PLECTIN 1.  
 FT PLECTIN 2.  
 FT PLECTIN 3.  
 FT PLECTIN 4.  
 FT PLECTIN 5.  
 FT PLECTIN 6.  
 FT PLECTIN 7.  
 FT PLECTIN 8.  
 FT PLECTIN 9.  
 FT PLECTIN 10.  
 FT PLECTIN 11.  
 FT PLECTIN 12.  
 FT PLECTIN 13.  
 FT PLECTIN 14.  
 FT PLECTIN 15.  
 FT PLECTIN 16.  
 FT PLECTIN 17.  
 FT PLECTIN 18.  
 FT PLECTIN 19.  
 FT PLECTIN 20.  
 FT PLECTIN 21.  
 FT PLECTIN 22.  
 FT PLECTIN 23.  
 FT PLECTIN 24.  
 FT PLECTIN 25.  
 FT PLECTIN 26.  
 FT PLECTIN 27.  
 FT PLECTIN 28.  
 FT PLECTIN 29.  
 FT PLECTIN 30.

Query Match	Best Local Similarity	6.2%	Score 113.5	DB 1	Length 4684
Matches 71	Conservative 63	Mismatches 176	Indels 51	Gaps 8	
14 AEOFRQKAL	-----RPAVIPREV	-----YQFDEMEASHLPYIEQSQMOL	55		
2123 AEERQKSLAAEEBAARQKALAEVERLKAKEBARSLRERAEQESARQLOLAQEAQK	2182				
56 AEOFRQKALFAASRGNHVSPTHYVESAPARAIIPREVLRREFDEMEAFASRGNHVS	115				
2183 RLQAEKKHAAHRAVQK	-----EQLOQTLQOQSVLDRIRGEAFA	-----RRAAEKEE	2232		
116 AHYVESDASQAAPYIEQAOVIAGOFKEKVALFA	-----SRGNHDSPTHYVESDAKPAIIPDR	174			
2233 ARVQREERAQSRQVEEAERLKQSAEBQAOQAQAAEKLKREABQEAAPRAQAEQA	2292				
175 EVLVRREFMEBSQHLPIYIEQSMALAEQFKQKALAFASRGNHVAATHYVTESDAKPAI	234				
2293 ALRQQAADAEAEKKKFAEQTLRQKAOVEOELTLRLQ	-----LEETDHOXNL	2342			
235 DKEVLVQOYDEMEAFASRGNHVAATHYVTESDASASLPYMETRAIAGOFKEKVALFAS	294				
2343 -----DELOQLKAEATEAARQORQVEEELFSVVOHELSKLARIEAERLALIL	2393				
295 RGNHVSPTHYVESAPARAIIPREVLRREFDEMEAFASRGNHVS	115				
2394 RDKNTOGRFLQDEEAQKMQVAEEAARLSVAQEAARLRLQLAEBDLAQORALAEKYLKEXM	2453				
352 Q 352					

```

DB      2454 Q 2454

RESULT 15
PLE1 RAT STANDARD; PRT; 4687 AA.
ID PLE1 RAT
AC P30427: 008879; 008880; 008881;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plectin 1 (PLTN) (PCN).
GN PLECL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Glial tumor;
RX MEDLINE=91268156; PubMed=2050743;
RA Wiche G., Becker B., Lubert R., Weltzer G., Castanon M.J.,
RA Hauptmann R., Stratawa C., Stewart M.;
RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
RT chain with a three-domain structure based on a central alpha-helical
RT coiled coil."
RL J. Cell Biol. 114:83-99(1991).
RN (2)
RP REVISIONS.
RC TISSUE=Glial tumor;
RX MEDLINE=96210632; PubMed=8633055;
RA Liu C.-G., Maerker C., Castanon M.J., Hauptmann R., Wiche G.;
RT "Human plectin: organization of the gene, sequence analysis, and
RT chromosome localization (8q24)."
RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN (3)
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4), AND TISSUE
RP SPECIFICITY.
RC TISSUE=Glial tumor;
RX MEDLINE=97321050; PubMed=9177781;
RA Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
RA Wiche G.;
RT "Plectin transcript diversity: identification and tissue distribution
RT of variants with distinct first coding exons and rodless isoforms.";
RL Genomics 42:115-125(1997).
CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=P30427-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P30427-2; Sequence=VSP_005050;
CC Name=3;
CC IsoId=P30427-3; Sequence=VSP_005051;
CC Name=4;
CC IsoId=P30427-4; Sequence=VSP_005052;
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
CC SKELETAL MUSCLE AND LOWEST IN THYMUS.
CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B, WHEREAS BOTH THE N-
CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC -1- PTM: PHOSPHORYLATED BY CDC3, REGULATES DISSOCIATION FROM
CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC -1- SIMILARITY: Contains 1 actin-binding domain.
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -1- SIMILARITY: Contains 3 plectin repeats.
CC -1- SIMILARITY: Contains 4 spectrin repeats.
CC -1- SIMILARITY: Belongs to the plectin or cytolinker family.
CC -1- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 13:06:36 ; Search time 41 Seconds

(without alignments)  
825.643 Million cell updates/sec

Title: US-09-491-146A-52

Perfect score: 1818

Sequence: 1 AAHPLYEQGMHLAEQFKOK.....KALIEEGQMAEMLSKIQ 352

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

11 number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	263	14.5	3010	1	S18030 genome polyprotein
2	260	14.3	3010	1	GNMVCU genome polyprotein
3	259	14.2	3011	1	S40770 genome polyprotein
4	258	14.2	3010	1	GNMVTW genome polyprotein
5	256	14.1	492	2	PS0326 polyprotein - hepa
6	256	14.1	3010	1	GNMVTW genome polyprotein
7	256	14.1	3010	1	A45573 genome polyprotein
8	254	14.0	3011	1	GNMVC3 genome polyprotein
9	247	13.6	3011	1	GNMVC3 genome polyprotein
10	207	11.4	716	2	J01366 polyprotein - hepa
11	203.5	11.2	3033	1	J01303 genome polyprotein
12	201.5	11.1	3033	1	GNMVT8 genome polyprotein
13	195	10.7	3014	1	UC5620 genome polyprotein
14	188.5	10.4	876	1	PC2219 polypeptide - hepa
15	188	10.3	125	2	S35629 hypothetical prote
16	176.5	9.7	209	2	PC1306 genome polyprotein
17	173.5	9.5	142	2	PC1307 genome polyprotein
18	164.5	8.5	411	2	S47436 flagellar antigen
19	121.5	6.7	911	2	UC7186 alpha-actinin-4 -
20	114.5	6.3	1621	2	A62255 hypothetical prote
21	114.5	6.3	5327	2	T13564 microtubule-associ
22	113.5	6.2	1224	2	T14007 microtubule-associ
23	113.5	6.2	1262	2	T22523 hypothetical prote
24	113.5	6.2	1778	2	AF1116 internalin protein
25	113.5	6.2	4574	2	G02520 plectin - human
26	113.5	6.2	4684	2	A59404 plectin (imported)
27	113.5	6.2	5170	2	T15348 hypothetical prote
28	113	6.2	7962	2	T18346 elastic titin - hu
29	112.5	6.2	1385	2	T21706 hypothetical prote

30	112.5	6.2	1929	2	T21559 hypothetical prote
31	112	6.2	1785	2	T21558 hypothetical prote
32	110.5	6.1	506	2	S47439 I2 protein - trypa
33	109.5	6.0	1948	2	S00485 gene 11-1 protein
34	109.5	6.0	4687	1	A39638 plectin - rat
35	107.5	5.9	1733	1	RNBX2L DNA-directed RNA p
36	107.5	5.9	5762	2	A41819 prolone-rich pepi
37	107	5.9	910	1	S73361 dnaJ homolog prote
38	107	5.9	2957	2	T33152 hypothetical prote
39	106.5	5.9	723	2	H85092 hypothetical prote
40	106.5	5.9	1145	2	T33606 hypothetical prote
41	106.5	5.9	1736	2	A47747 tight junction pro
42	106	5.8	948	2	S75991 hypothetical prote
43	106	5.8	2748	2	S57976 nuclear migration
44	104.5	5.7	327	2	S31781 HRP1 protein - ma
45	104.5	5.7	419	2	G75062 probable flagella-

#### ALIGNMENTS

##### RESULT 1

S18030 genome polyprotein - hepatitis C virus (isolate JKI)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstr) protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

A:Variety: isolate JKI

C:Date: 19-May-2000 #sequence revision 19-May-2000 #text\_change 23-Mar-2001

C:Accession: S18030; S33570; A48332; S18029

R:Honda, M.; Kaneko, S.; Maseh, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single i

A:Reference number: S18028

A:Accession: S18030

A:Molecule type: genomic RNA

A:Residues: 1-310 <HON>

A:Cross-references: EMBL:X61596; NID:959478; PIRN:CAA4793.1; PID:959479

A:Experimental source: isolate JKI from an individual

R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.

Arch. Virol. 128, 163-169, 1993

A:Title: Sequence analysis of putative structural regions of hepatitis C virus isola

A:Reference number: A48332; MUID:91119270; PMID:8380322

A:Accession: S33570

A:Molecule type: genomic RNA

A:Residues: 1-547, 'T', '549-621', 'V', '623-624', 'S', '626-652', 'DL', '655-761', 'T', '763-782' <HOW>

A:Cross-references: EMBL:X61591

A:Note: this sequence is inconsistent with the nucleotide translation

as Trp, and Trp for residue 771 as Ser

A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; s

F:2-115/Product: capsid protein C #status predicted <CC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <ME>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus status predicted <NS3>

F:1120-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEHH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match 14.5%; Score 263; DB 1; Length 3010;  
Best Local Similarity 32.5%; Pred. No. 3.1e-09;  
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 11;







Db 1746 -----AAP-----VSESKRALETFPAKHMNMNISGIQYLAGSLTPGNPAIASLMAF 1793  
QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVESPDAKPAI-----170  
Db 1794 TASITSPLTOSTLFTNLIGGVAAQLAPSAASAFVAGIAGAAGVSGIGKVLVDILA 1853  
QY 171 -----IPDREVLVREFDEMSESHL--PYIEQGMML-----AE 201  
Db 1854 GYGAGVAGALVAFKVMSEMPSTEDLVNLLPALISPGALVGVCAILRRHVGPGEAV 1913  
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228  
Db 1914 QMNRRLIAFASRGNHVSPTHYVESPDA 1940

RESULT 7  
A45573  
Genome polyprotein - hepatitis C virus (strain JT)  
N Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
elN NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C; ecies: hepatitis C virus  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C/Accession: A45573  
R/Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,  
virus Res. 23, 39-53, 1992  
A/Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S  
A/Reference number: A45573; MUID:92295714; PMID:1318627  
A/Accession: A45573  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-3010 <TM>  
A/Cross-references: GB:D11168; GB:D01171; NID:9221612; PIDN:BA01943.1; PID:9221613  
A/Experimental source: HCV-JT  
A/Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:P.106207)  
C/Suprafamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; seiri  
F/116-191/Product: capsid protein C #status predicted <CPC>  
F/112-115/Product: major envelope protein M #status predicted <EPM>  
F/132-389/Product: major envelope protein E #status predicted <ME>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepatitis virus #status predicted <NS3>  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1319/Region: nucleotide-binding motif B  
F/116-1319/Region: DEKH motif  
F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F/1616-1862/Product: nonstructural protein NS4b #status predicted <N4B>  
F/1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 14.1%; Score 256; DB 1; Length 3010;  
Best Local Similarity 31.5%; Pred. NO. 9e-09;  
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;

QY 23 RPAAVDPREVLYOEFDMEEM--ASHLPYIEQGMQLAEQFKOKALAF---ASRGNHVSPTHY 78  
Db 1691 RPAAVDPREVLYREFDEMEECSHLPYIEQGMQLAEQFKOKALGLQTAATKQ-----1742  
QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSPTHYVESPDAQAAPYIEQ-----133  
Db 1743 ---AEEAAPVSESR-----WRALFAFMAGHMNFIISGIOYLAGSLTPGNPAIASLMAF 1793  
QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVESPDAKPAI-----170  
Db 1794 TASITSPLTOSTLFTNLIGGVAAQLAPSAASAFVAGIAGAAGVSGIGKVLVDILA 1853  
QY 171 -----IPDREVLVREFDEMSESHL--PYIEQGMML-----AE 201  
Db 1854 GYGAGVAGALVAFKVMSEMPSTEDLVNLLPALISPGALVGVCAILRRHVGPGEAV 1913  
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228  
Db 1914 QMNRRLIAFASRGNHVSPTHYVESPDA 1940

RESULT 8  
GNWVC3  
Genome polyprotein - hepatitis C virus (strain HCV-1)  
N/Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonst  
C/Species: hepatitis C virus  
C/Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 19-Jan-2001  
C/Accession: A39166; P00403  
R/Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.  
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A/Title: Genetic organization and diversity of the hepatitis C virus.  
A/Reference number: A39166; MUID:91172826; PMID:1848704  
A/Accession: A39166  
A/Molecule type: mRNA  
A/Residues: 1-3011 <CHO>  
A/Cross-references: GB:M62321; NID:9329873; PIDN:AAA5676.1; PID:9329874  
R/Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap  
J. Gen. Virol. 73, 1131-1141, 1992  
A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship  
A/Reference number: P00393; MUID:92268871; PMID:1316939  
A/Accession: P00403  
A/Molecule type: genomic RNA  
A/Residues: 1577-1633 <CH>  
A/Cross-references: DDBJ:D10128  
A/Experimental source: isolates E-b16  
A/Accession: P00404  
A/Status: preliminary  
A/Molecule type: genomic RNA  
A/Residues: 1577-1633 <CH>  
A/Experimental source: isolates E-b17  
A/Suprafamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru  
F/116-191/Product: capsid protein C #status predicted <CPC>  
F/112-115/Product: major envelope protein M #status predicted <EPM>  
F/132-389/Product: major envelope protein E #status predicted <ME>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: nucleotide-binding motif A (P-loop)  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1319/Region: nucleotide-binding motif B  
F/116-1319/Region: DEKH motif  
F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F/1616-1862/Product: nonstructural protein NS4b #status predicted <N4B>  
F/1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,20

Query Match 14.0%; Score 254; DB 1; Length 3011;  
Best Local Similarity 31.8%; Pred. No. 1.2e-08;  
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY 23 RPAAVDPREVLYOEFDMEEM--HLPYIEQGMQLAEQFKOKALAF---ASRGNHVSPTHY 78  
Db 1691 RPAAVDPREVLYREFDEMEECSHLPYIEQGMQLAEQFKOKALGLQTAATKQAEV-----1745  
QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSPTHYVESPDAQAAPYIEQ-----133  
Db 1746 -----IAPAV-----QTNQKLETFPAKHMNMNISGIQYLAGSLTPGNPAIASLMAF 1793  
QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVESPDAKPAI-----170  
Db 1794 TAAVTSPLTOSTLFTNLIGGVAAQLAPCAAFATFVAGIAGAAGVSGIGKVLVDILA 1853  
QY 171 -----IPDREVLVREFDEMSESHL--PYIEQGMML-----AE 201  
Db 1854 GYGAGVAGALVAFKVMSEMPSTEDLVNLLPALISPGALVGVCAILRRHVGPGEAV 1913  
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228  
Db 1914 QMNRRLIAFASRGNHVSPTHYVESPDA 1940

## RESULT 9

GNMVCH

genome polypeptide - hepatitis C virus (strain H)

N.Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C.Species: hepatitis C virus

A.Note: host Homo sapiens (man)

C.Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001

C.Accession: A36814; A41546

R.Inchoupe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Submitted to GenBank, July 1992

A.Description: Genomic structure of the human prototype strain H of hepatitis C virus; C

A.Reference number: A36814

A.Accession: A36814

A.Molecule type: genomic RNA

A.Residues: 1-3011 &lt;INC&gt;

A.Cross-references: GB:M67463, NID:G229737, PIDN:AAA5534.1, PID:G229738

R.Inchoupe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991

Title: Genomic structure of the human prototype strain H of hepatitis C virus; compari

Reference number: A41546; MUID:92052256; PMID:1658800

Contents: annotation

A.Note: neither amino acid nor nucleotide sequence is given

C.Superfamily: hepatitis C virus genome polypeptide

C.Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructura

F.1-115/Product: capsid protein C #status predicted &lt;CPC&gt;

F.116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F.192-389/Product: major envelope protein E #status predicted &lt;ME&gt;

F.390-723/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F.730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F.1007-1615/Product: hepacivirin #status predicted &lt;NS3&gt;

F.1320-1337/Region: nucleotide-binding motif A (P-loop)

F.1312-1317/Region: nucleotide-binding motif B

F.1316-1319/Region: DEKH motif

F.1616-1662/Product: nonstructural protein NS4a #status predicted &lt;NS4&gt;

F.1867-2013/Product: nonstructural protein NS4b #status predicted &lt;NS4b&gt;

F.2014-3011/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F.196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match

Best Local Similarity 13.6%; Score 247; DB 1; Length 3011;

Matches 86; Conservative 26; Mismatches 77; Indels 78; Gaps 11;

23 RPAVLPDREVLVYREFDEMEAS-HLPYIEOGMLAEQFOKXALAF---ASRGNHVSPTHY 78

Db 1691 KPAIIPDREVLVYREFDEMECSOHLPIYIEOGMLAEQFOKXALGLLQTSRAAEV----- 1745

QV 79 VESDARPAIIPDREVLVYREFDEMEASR-GNHVSPTHYVPSDASQAFYIEO--- 133

L 1746 -----ITPAV-----QTNMOKLEVFMAKHMNFISGIQYLAGLSTLPGNPAISLMAF 1793

QY 134 -AOVIAHOFKEKYLAF-----ASRGHDSPTHYVPSDAKPAI----- 170

Db 1794 TAAVVSPLTTGQTLLENILGWVAAOLAPGAATAFVAGLGLAGLDSVGLKVLVDILA 1853

QY 171 -----IPDREVLVYREFDEMECSOHLPIYI-----EQGMMLAE----- 201

Db 1854 GYGAGVAGALVAFKISGEVSTEDLVNLPAILSPGALAVGVFASIRRVGREGAV 1913

QY 202 QFOKALAFASRGNHVAPTHYTESDA 228

Db 1914 QMMNRLIAFASRGNHVSPTHYVESDA 1940

## RESULT 10

JQ1366

polypeptide - hepatitis C virus (French isolate) (fragments)

C.Species: hepatitis C virus

C.Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 17-Nov-2000

C.Accession: JQ1366

R.Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A&gt;Title: Partial nucleotide sequence analysis of a French hepatitis C virus; implication

A.Reference number: JQ1366; MUID:92013977; PMID:1655961

A.Accession: JQ1366

A.Molecule type: genomic RNA

A.Residues: 1-716 &lt;KRE&gt;

C.Superfamily: hepatitis C virus genome polypeptide

C.Keywords: glycoprotein; polypeptide

F.84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Aen) (covalent)

Query Match

Best Local Similarity 11.4%; Score 207; DB 2; Length 716;

Matches 49; Conservative 4; Mismatches 16; Indels 16; Gaps 3;

QY 167 KPAIIPDREVLVYREFDEME-SOHLPIYIEOGMLAEQFOKXALAFASRGNHVAPTHYTE 225

Db 640 KPAIIPDREVLVYREFDEMECSOHLPIYIEOGMLAEQFOKXALGLLQ-----TR 688

QY 226 SDKAPALVPDKVLYQYDEMEAF 250

Db 689 SROAEVIRPAVOTMQR-----EAF 709

## RESULT 11

JQ1303

genome polypeptide - hepatitis C virus (isolate HC-06)

N.Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonst

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C.Species: hepatitis C virus

C.Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000

C.Accession: JQ1303

R.Okamoto, H.; Okada, S.; Sugiyama, Y.; Kura, K.; Itzuka, H.; Machida, A.; Miyakaw

J. Gen. Virol. 72, 2697-2704, 1991

A&gt;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from

A.Reference number: JQ1303; MUID:92044440; PMID:1658196

A.Accession: JQ1303

A.Molecule type: genomic RNA

A.Residues: 1-3033 &lt;OKA&gt;

A.Cross-references: GB:D00944; NID:G221650; PIDN:BA00792.1; PID:G221651

A.Experimental source: isolate HC-06 from a Japanese individual

C.Superfamily: hepatitis C virus genome polypeptide

C.Keywords: ATP; glycoprotein; hydrolase; P-loop; polypeptide; serine proteinase; c

F.2-115/Product: capsid protein C #status predicted &lt;CPC&gt;

F.116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F.192-389/Product: major envelope protein E #status predicted &lt;ME&gt;

F.390-733/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F.734-1010/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F.1011-1619/Product: hepacivirin #status predicted &lt;NS3&gt;

F.1316-1321/Region: nucleotide-binding motif B

F.1320-1323/Region: DEKH motif

F.1620-1866/Product: nonstructural protein NS4a #status predicted &lt;NS4&gt;

F.1867-2017/Product: nonstructural protein NS4b #status predicted &lt;NS4b&gt;

F.2018-3033/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F.196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20;

Query Match

Best Local Similarity 11.2%; Score 203.5; DB 1; Length 3033;

Matches 60; Conservative 10; Mismatches 37; Indels 29; Gaps 5;

QY 229 KPAVLPDKVLYQYDEMEAFASRGNHVAPTHYVESDAS---ASLPYMETRAIAQOF 285

Db 1619 KPTLVGPPLLYR-----LGSVTNEVLTTHPTKXIAICMOMADEVMSTWLACGV 1670

QY 286 KEXVIAFASRGN-----HVSPTHYVPESEPOVVYPPDKELLYEAFDEME-ASKAAL 336

Db 1671 LAAYAYCLATGCACCIIRLHVNR-----AVVAPDKVLYEAFDEMECSRAAL 1721

QY 337 IEEGORMAEMLSKIQ 352

Db 1722 IEEGRIAMLSKIQ 1737

## RESULT 12

GNMV08

genome polypeptide - hepatitis C virus (strain HC-08)

N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C>Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C/Accession: A40250; P00397; P00559  
C/Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;  
Virology 188, 331-341, 1992  
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep  
A:Reference number: A40250; MUID:9220323; PMID:1314459  
A/Accession: A40250  
A/Molecule type: genomic RNA  
A/Residues: 1-3033 <OKA>  
A/Cross-references: GB:D10988; GB:D01221; NID:G221608; PIDN:BAA01761.1; PID:G221609  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EMP>  
F:192-389/Product: major envelope protein E #status predicted <ME>  
J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
A:Reference number: P00393; MUID:92268871; PMID:1316939  
A/Accession: P00397  
A/Molecule type: genomic RNA  
A/Residues: 2678-2754 <CHA>  
A/Cross-references: DDBJ:D10134  
A:Experimental source: isolate E-b12  
R.Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnd  
Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan.  
A:Reference number: P00554; MUID:92068204; PMID:1720309  
A/Accession: P00559  
A/Molecule type: mRNA  
A/Residues: 2678-2729 <KAT>  
A/Cross-references: GB:D10562; GB:D90518; NID:G221523; PIDN:BAA01418.1; PID:G221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EMP>  
F:192-389/Product: major envelope protein E #status predicted <ME>  
F:794-1010/Product: nonstructural protein NS1 #status predicted <NS1>  
F:1011-1619/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1234-1241/Region: nucleotide-binding motif A (P-loop)  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 11.1%; Score 201.5; DB 1; Length 3033;  
Best Local Similarity 46.2%; Pred. No. 3.9e-05;  
atches 60; Conservative 13; Mismatches 40; Indels 17; Gaps 5;

QY 229 KPALVDKELVYQYDEMEAFASRGHVAFTHYVESDAS--ASLPYDETRAIAGQF 285  
DB 1619 KPLTGPPTPLLYR-----LGAVTNEVLTTHPVKRYATCQADLEIMTSSWVLGCV 1670

QY 286 KEVYLFASRGHVS--PRHYPESEPOVVTPDKETLVEAFPEME--ASKALIEGOR 342  
DB 1671 LAVAAYCATGCTGCTIIGRLHND--RVVAPDKETLVEAFPEMECSKALIEGOR 1727

QY 343 MAEMLSKIQ 352  
DB 1728 MAEMLSKIQ 1737

RESULT 13  
JC5620  
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C/Accession: JC5620  
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomi  
A:Reference number: JC5620; MUID:97366593; PMID:9223423  
A/Accession: JC5620  
A/Molecule type: mRNA  
A/Residues: 1-3014 <CHA>  
A/Cross-references: GB:Y13184  
A:Experimental source: genotype 5a, which predominates in South Africa  
A>Note: the translation of the nucleotide sequence is not complete in this paper  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EMP>  
F:192-389/Product: major envelope protein E #status predicted <ME>  
F:384-408/Region: hypervariable #status predicted  
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1008-1616/Product: hepatitisin #status predicted <NS3>  
F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
F:1313-1318/Region: nucleotide-binding motif B  
F:1317-1320/Region: DEXH motif  
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2210-2249/Region: interfeon sensitivity determining #status predicted

Query Match 10.7%; Score 195; DB 1; Length 3014;  
Best Local Similarity 26.6%; Pred. No. 0.00011;  
Matches 80; Conservative 40; Mismatches 127; Indels 54; Gaps 10;

QY 23 RPAVLPDEPVLYOEDEMEAS-HLPYIEGQQLAEQFKKALAFASRGHVSPTHYVE 81  
DB 1692 RAIPTRDREVLVYQYDEMECSASLPYDEARAIAGQFEKYLIGTAGQAKET----- 1746

QY 82 SDARPAIIPDREVLHREFDEMEAFASR-GNHVSFAHYVESDASQAAPYIEQAQVIAHQ 140  
DB 1747 -LKRPAAT-----SMSSKAEQFPAKHMNVSIGVLAGSTLPGNPVAV----- 1788

QY 141 FEKYLAFASRGHNDSPHYVESDAPKPAIIPDREVLHREFDEMEESGHLPIEQQMILA 200  
DB 1789 -ATLMSFTAAVTSPLTTH-----QTLLFNILGQWVASQIAP-----PTAA 1827

QY 201 EGFKKALAFASRGHVAFTHYVE-----SDAKPALVPDEKLYQYDEMEAFASRG 254  
DB 1828 TAFVSGMAAGAVNG-IGLGRVLIDILAGYGVGVALVAFK-INCGRPTAAETLVNLLP 1885

QY 255 NHVAFTHYVESDASASLPYDETRAIAGQFEKYLAFASRGHVSPTHYVESEPOVV 314  
DB 1886 SLTCGALVVGICAAVLRHHIGPEGAVQMMNRILAFASRGHNSPTHYVEPTASAKV 1945

QY 315 T 315  
DB 1946 T 1946

RESULT 14  
PC2219  
N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A )  
C:Species: hepatitis C virus  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 17-Nov-2000  
C/Accession: PC2219  
R:Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.  
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994  
A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of  
A:Reference number: PC2219; MUID:94338342; PMID:7520237  
A/Accession: PC2219  
A/Molecule type: mRNA  
A/Residues: 1-876 <STU>  
A/Cross-references: GB:L29577; GB:L29578; GB:L29579  
A:Experimental source: serum  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: glycoprotein  
F:1-191/Product: core #status predicted <COB>

F:68-78/Region: variable  
F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>  
F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>  
F:248-338/Region: E2  
F:339-411/Region: NS1 (amino end)  
F:412-783/Product: NS3 #status predicted <NSR>  
F:784-837/Product: NS4A #status predicted <NSA>  
F:838-876/Product: NS4B #status predicted <NSB>  
F:281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.4%; Score 188.5; DB 2; Length 876;  
Best Local Similarity 59.1%; Pred. No. 5,7e-05;  
Matches 39; Conservative 4; Mismatches 4; Indels 19; Gaps 1;

QY 229 KPALVPDKKVLVYQYDEMEEPFASRCGNHVAPTHYVVESDASLPPYMDETRAIAGQFKK 288  
DB 817 KPALIPDKKVLVYQYDEMEEPFASRCGNHVAPTHYVVESDASLPPYMDETRAIAGQFKK 288

QY 289 VLAFAAS 294  
858 VLGFIS 863

## RESULT 15

S35629  
hypothetical protein - hepatitis C virus

C:Species: hepatitis C virus

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Nov-2000

C:Accession: S35629

R:Sarashina, T.; Sakurai, T.; Watanabe, Y.; Kashima, K.; Suzuki, T.; Chiba, J.; Kita, Y.

Nucleic Acids Res. 21, 1037, 1993

A:Title: Nucleotide sequence of the hepatitis C virus genome from a patient negative for

A:Reference number: S35629; MID:93197128; PMID:8383835

A:Accession: S35629

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-125 <SAR>

A:Cross-references: EMBL:DI1353; NID:G221625; PIDN:BA01956.1; PID:G221626

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992

C:Superfamily: hepatitis C virus genome polypeptide

## Query Match

Best Local Similarity 10.3%; Score 188; DB 2; Length 125;

Matches 49; Conservative 10; Mismatches 32; Indels 6; Gaps 3;

QY 259 PTHYVVESDASLPPYMDETRAIAGQFKKVLAFASRCGNHVS--PRHYVESEPPQVVTVP 316  
DB 2 PTHYVVESDASLPPYMDETRAIAGQFKKVLAFASRCGNHVS--PRHYVESEPPQVVTVP 316

QY 317 DKEILYEAPEPMEE-ASKAALIEGORMEMLKSKIO 352  
DB 59 DKEILYEAPEPMEE-ASKAALIEGORMEMLKSKIO 352

Search completed: July 25, 2003, 13:12:47  
Job time : 43 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 13:12:06 / Search time 52 Seconds  
(without alignments)  
803.912 Million cell updates/sec

Title: US-09-491-146a-52

Perfect score: 1818  
Sequence: 1 AAH1PYLEQGMHLAEQFKOK.....KAALEEGQMAEWLKSQIQ 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES:

Result No.	Score	Query Match	Length	ID	Description
1	512.5	28.2	829	10 US-09-881-239-5	Sequence 5, Appl1
2	512.5	28.2	1099	10 US-09-881-654-4	Sequence 4, Appl1
3	256	14.1	2201	14 US-10-085-476-2	Sequence 2, Appl1
4	255	14.0	2985	15 US-10-259-275-40	Sequence 40, Appl1
5	255	14.0	3011	9 US-09-742-659-4	Sequence 4, Appl1
6	255	14.0	3011	11 US-09-891-894-3	Sequence 3, Appl1
7	255	14.0	3012	10 US-09-238-076-2	Sequence 2, Appl1
8	255	14.0	3012	11 US-09-995-937-2	Sequence 2, Appl1
9	255	14.0	3012	11 US-09-917-563-2	Sequence 2, Appl1
10	254	14.0	2894	10 US-09-941-611-23	Sequence 23, Appl1
11	254	14.0	2894	15 US-10-044-995-23	Sequence 23, Appl1
12	254	14.0	3011	9 US-09-916-359-2	Sequence 2, Appl1
13	251	13.8	1985	15 US-10-259-275-42	Sequence 42, Appl1
14	251	13.8	3011	16 US-10-232-643-6	Sequence 6, Appl1
15	250	13.8	3011	10 US-09-238-076-20	Sequence 20, Appl1

16	250	13.8	3011	10 US-09-952-572-9	Sequence 9, Appl1
17	250	13.8	3011	10 US-09-747-419-20	Sequence 20, Appl1
18	250	13.8	3011	11 US-09-995-937-20	Sequence 20, Appl1
19	250	13.8	3011	11 US-09-917-563-20	Sequence 20, Appl1
20	250	13.8	3011	15 US-10-259-275-20	Sequence 20, Appl1
21	249	13.7	2201	14 US-10-029-907-3	Sequence 3, Appl1
22	247	13.6	3011	10 US-09-929-955-1	Sequence 1, Appl1
23	247	13.6	3011	14 US-10-104-966-1	Sequence 1, Appl1
24	205	11.3	79	9 US-09-758-308-3	Sequence 3, Appl1
25	200	11.0	1692	11 US-09-919-901-4	Sequence 4, Appl1
26	200	11.0	1692	11 US-09-919-901-11	Sequence 11, Appl1
27	200	11.0	1692	11 US-09-919-901-18	Sequence 18, Appl1
28	200	11.0	2307	11 US-09-919-901-2	Sequence 2, Appl1
29	200	11.0	2307	11 US-09-919-901-9	Sequence 9, Appl1
30	200	11.0	2307	11 US-09-919-901-16	Sequence 16, Appl1
31	197.5	10.9	54	10 US-09-821-397-8	Sequence 8, Appl1
32	192.5	10.6	484	11 US-09-899-046-198	Sequence 198, Appl1
33	192.5	10.6	484	11 US-09-878-281-198	Sequence 198, Appl1
34	188.5	10.4	481	11 US-09-899-046-270	Sequence 270, Appl1
35	188.5	10.4	481	11 US-09-878-281-270	Sequence 270, Appl1
36	188.5	10.4	484	11 US-09-899-046-200	Sequence 200, Appl1
37	188.5	10.4	484	11 US-09-878-281-200	Sequence 200, Appl1
38	182.5	10.0	128	11 US-09-899-046-62	Sequence 62, Appl1
39	182.5	10.0	128	11 US-09-878-281-62	Sequence 62, Appl1
40	179.5	9.9	128	11 US-09-899-046-60	Sequence 60, Appl1
41	179.5	9.9	128	11 US-09-878-281-60	Sequence 60, Appl1
42	176.5	9.7	133	11 US-09-899-046-38	Sequence 38, Appl1
43	176.5	9.7	133	11 US-09-899-046-40	Sequence 40, Appl1
44	176.5	9.7	133	11 US-09-878-281-38	Sequence 38, Appl1
45	176.5	9.7	133	11 US-09-878-281-40	Sequence 40, Appl1

## ALIGNMENTS

RESULT 1	US-09-881-239-5
Sequence 5, Application US/09881239	
Publication No. US20020192639A1	
GENERAL INFORMATION:	
APPLICANT: CHIEN, David Y.	
APPLICANT: ARCANDEL, Philipp	
APPLICANT: TANDESKE, Laura	
APPLICANT: GEORGE-NASCIMENTO, Carlos	
APPLICANT: COIT, Doris	
APPLICANT: MEDINA-SELBY, Angelica	
TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY	
FILE REFERENCE: 2302-16073 / PP16073.003	
CURRENT APPLICATION NUMBER: US/09/881,239	
CURRENT FILING DATE: 2001-06-14	
NUMBER OF SEQ ID NOS: 8	
SOFTWARE: PatentIn Ver. 2.0	
SEQ ID NO 5	
LENGTH: 829	
TYPE: PRT	
ORGANISM: Artificial Sequence	
FEATURE:	
OTHER INFORMATION: Description of Artificial Sequence: MEFA 12	
US-09-881-239-5	
Query Match	28.2%; Score 512.5; DB 10; Length 829;
Best Local Similarity	62.8%; Pred. No. 8.3e-35;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;	
QY 167 KPAIIPDREVLVEPEFDEME-SCHLPYIEGGMHLAEQFKOKALAFASRGHVAPTHVTE 225	
Db 397 KPAIIPDREVLVEPEFDEME-SCHLPYIEGGMHLAEQFKOKALG-L-SRG----- 444	
QY 226 SDKAPALVPKVELVYQYIDEMEPAFASRGHVAPTHVVSASISPIYDETRALAGOF 285	
Db 445 --GKPAIVPKEVLVYQYIDEMEPAFASRGHVAPTHVVSASISPIYDETRALAGOF 483	
QY 286 KENVLAFASGNHVSRRHYVPESEPOVVTPDKETILYEAFDEME-ASKKALIEEGQMA 344	

Db 484 KEKVGLI-----DNDQVVVTPDKELLYEAFDEMECASALIEGQRNA 529  
QY 345 EMLKSKIQ 352  
Db 530 EMLKSKIQ 537

## RESULT 2

US-09-881-654-4  
; Sequence 4, Application US/09881654  
; Patent No. US2002014685A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIEN, David Y.  
; APPLICANT: ARCANGEL, Phillip  
; APPLICANT: TRANDSKE, Laura  
; APPLICANT: GEORGE-NASCIMENTO, Carlos  
; APPLICANT: COLT, Doris  
; APPLICANT: MEDINA-SELBY, Angelica  
; TITLE OF INVENTION: IMMUNOSSAYS FOR ANTI-HCV ANTIBODIES  
; FILE REFERENCE: 2302-17039 / PPI7039, 002  
; CURRENT APPLICATION NUMBER: US/09/881,654  
; PRIOR FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: 60/212,082  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 60/280,811  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/280,867  
; PRIOR FILING DATE: 2001-04-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1099  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: MEPA 7.1  
US-09-881-654-4

Query Match 28.2%; Score 512.5; DB 10; Length 1099;  
Best Local Similarity 62.8%; Pred. No. 1.2e-34;  
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

QY 167 KEAIIIPDEVLVREFDEME--SOHLPYIEQGMMLAEQFOKALAFASRGNHVAPTHVYTE 225  
Db 701 KEAIIIPDEVLVREFDEMECSQHPIYIEQGMMLAEQFOKALG-SRG----- 748  
QY 226 SDAKALVDPKEVLVQOYDEMEAFASRGNHVAPTHVYVESDASALPYMDETRAIAGOF 285  
Db 749 --CKRAIVDPKEVLVQOYDEME-----CSQAAPIIEQAQVIAHQF 787  
QY 286 KEKVLAFASRGNHVSPRHVYPESEPOVVTPDKELLYEAFDEME--ASKALIEGQRNA 344  
Db 788 KEKVGLI-----DNDQVVVTPDKELLYEAFDEMECASALIEGQRNA 833  
QY 345 EMLKSKIQ 352  
Db 834 EMLKSKIQ 841

## RESULT 3

US-10-085-476-2  
; Sequence 2, Application US/10085476  
; Publication No. US20020164722A1  
; GENERAL INFORMATION:  
; APPLICANT: De Francesco, Raffaele  
; APPLICANT: Tomei, Lucia  
; APPLICANT: Behrens, Sven-Erik  
; TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE  
; TITLE OF INVENTION: RNA-DEPENDENT RNA POLYMERASE AND TERMINAL NUCLEOTIDYL  
; FILE REFERENCE: IT0002PCA  
; CURRENT APPLICATION NUMBER: US/10/085,476

; CURRENT FILING DATE: 2002-02-27  
; PRIOR APPLICATION NUMBER: 08/952,981  
; PRIOR FILING DATE: 1998-03-23  
; PRIOR APPLICATION NUMBER: PCT/IT96/00106  
; PRIOR FILING DATE: 1996-05-24  
; PRIOR APPLICATION NUMBER: RM95A000343  
; PRIOR FILING DATE: 1995-05-25  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: CDNA clone pCD (38-9.4)  
US-10-085-476-2

Query Match 14.1%; Score 256; DB 14; Length 2201;  
Best Local Similarity 31.1%; Pred. No. 1.1e-12;  
Matches 83; Conservative 31; Mismatches 75; Indels 76; Gaps 10;

QY 23 RPAVDPDEVLVREFDEME--ASHPIYIEQGMMLAEQFOKALAF--ASRGNHVSPTHY 78  
Db 882 RPAVDPDEVLVREFDEMECSHPIYIEQGMMLAEQFOKALGILQTRAQAEA----- 936  
QY 79 VPESDARPAIIPDEVLVREFDEMEAFASR-GNHVSPAHVVPESDASQAAPYIEQ----- 133  
Db 937 ----AAP-----VESKRALETFAKMMNFIISIQYLAGLSTLPGNPALASLMAF 984  
QY 134 -AQVIAHQFKEKVLAF-----ASRGNHSPTHVVPESDAKPAI----- 170  
Db 985 TASITSPITQSTLFLNLIGWVAQAAPASAFVAGAGAAVSGIGKVLVDILA 1044  
QY 171 -----IPDEVLVREFDEME--PYIEQGMML-----AE 201  
Db 1045 GYGAGVAGALVAFKMGEMSTEDLVNLPALISPGALVGVCAILRRHVPGEGAV 1104  
QY 202 QFOKALAFASRGNHVAPTHVYTESDA 228  
Db 1105 QMNRLLAFASRGNHVSPTHVVPESDA 1131

## RESULT 4

US-10-259-275-40  
; Sequence 40, Application US/10259275  
; Publication No. US20030125541A1  
; GENERAL INFORMATION:  
; APPLICANT: Lemon, Stanley M.  
; APPLICANT: Yi, Minkyung  
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE  
; FILE REFERENCE: 265,0007,0120  
; CURRENT APPLICATION NUMBER: US/10/259,275  
; PRIOR FILING DATE: 2003-01-13  
; PRIOR APPLICATION NUMBER: US 60/171,909  
; PRIOR FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: US 09/747,419  
; PRIOR FILING DATE: 2000-12-23  
; PRIOR APPLICATION NUMBER: US 60/325,236  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US 60/338,123  
; PRIOR FILING DATE: 2001-11-13  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 40  
; LENGTH: 2985  
; TYPE: PRT  
; ORGANISM: ARTIFICIAL  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence encoded by nucleotides 2077-11121 of SEQ  
US-10-259-275-40

Query Match 14.0%; Score 255; DB 15; Length 2985;  
Best Local Similarity 31.5%; Pred. No. 2.1e-12;  
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;







```

      TELECOMMUNICATION INFORMATION:  
        TELEPHONE: 314-727-5188  
        TELEFAX: 314-727-6092  
  
    INFORMATION FOR SEQ ID NO: 2:  
      SEQUENCE CHARACTERISTICS:  
        LENGTH: 3012 amino acids  
        TYPE: amino acid  
        STRANDEDNESS: single  
        TOPOLOGY: linear  
      MOLECULE TYPE: protein  
      HYPOTHEetical: NO  
      FRAGMENT TYPE: N-terminal  
      SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
us-09-917-563-2  
  
Query Match          14.0%, Score 255; DB 11; Length 3012;  
Best Local Similarity 32.2%; Pred. No. 2,1e-12;  
Matches 86; Conservative 26; Mismatches 77; Indels 78; Gaps 10,  
  
QY      23 RPAVIPPREVLYQGFDEMEENS-HLPIYEQGMLAEQFKOKALAF--ASRQHNVSPHY 78  
       :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||  
Db      1691 KPAIIPREVLYQGFDEMECSOHLPYIEQGMMLAEQFKOKALGLQTASQAEEV----- 1745  
QY      79 VPESDAPPAIPREVLHREFDEMEEAFASR-GNHVSFAHYVPESDAQAPYIEQ---- 133  
       ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||  
Db      1746 -----ITPAV-----OTNWKQKLEVFNAKMNNFISGIOTLAGISTLPGNPAIASLMAF 1793  
QY      134 -AQVIAHQFEKEVLAF-----ASRGNDSPTHYVPESDAKPAI----- 170  
Db      1794 TAAVTSLTTSQTLFLFNILSGWVAQLAARPAATAFVAGLAGAIGSVGKLVLDILA 1853  
QY      171 -----IPDEVLYREFDEMEESCOHL--PYIEQGMML-----AE 201  
Db      1854 GYGAGVAGALAAFFKIMSGEVPESTDVLPLPALSPCALVGVCAAILRRHVPGEGAV 1913  
QY      202 QFKOKALAFASRGNHVAPTHYVTESDA 228  
       |:|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||  
Db      1914 QMNMRLIAFASRGNHVSPTHYVPESDA 1940  
  
RESULT 10  
US-09-941-611-23  
Sequence 23, Application US/09941611  
Patent No. US20020106640A1  
GENERAL INFORMATION:  
APPLICANT: DELEYS, ROBERT J  
            POLETT, DIRK  
            MAERTENS, GEERT  
            VAN HEUVERSMUN, HUGO  
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF  
                ANTI-BODIES TO HEPATITIS C VIRUS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHYE P.C.  
STREET: 1100 NORTH GLEEBO ROAD  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/941,611  
FILING DATE: 30-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/391,671  
FILING DATE: 1995-02-21  
APPLICATION NUMBER: WO PCT/EP91/02409  
FILING DATE: 13-DEC-1991
```

APPLICATION NUMBER: EP 90124241.2  
FILING DATE: 14-DEC-1990

ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B.J.  
REGISTRATION NUMBER: 36,663  
REFERENCE//DOCKET NUMBER: 1487-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 7038164000  
TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2894 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEetical: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 23:

US-09-941-611-23

Query Match            14.0%; Score 254; DB 10; Length 2894;  
Best Local Similarity 31.8%; Pred. No. 2,4e-12;  
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY     23 RPAVIPREVLYOEFDMEEAS-HLPYEQGMOLAEQFKOKALAF---ASRQNHSPTHY 78  
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db     1691 KPAPIPREVLYREFDEMECSQHLPYIEQGMLAEQFKOKALLGQTASRAQEV----- 1745

QY     79 VPESDAPRALIIPREVLHREFDEMEEAFASR-GNHVSAPHYVPESDSQAAPYEQ---- 133  
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db     1746 ----IAPVA-----QTNWOKLETFWAKHMNFISGIQLAGLSTLPGNPAIASLMAF 1793

QY     134 -AQVIAHQFEKYLAF-----ASGRNHSPTHYVPESDAKPPI----- 170  
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db     1794 TAAVTSLTTSQTLPLFIILGWVAALQAACGAATAFVGAGLAGAIALSVGLGYLDILA 1853

QY     171 -----IPDEVLYREFDEMEESQHL--PYIEQGMUL-----AE 201  
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db     1854 GYGAVGAGVALVAFKMGGEVPTEDLVNLPLPALSPGALVVGVCAAILRRHVGPGECAV 1913

QY     202 QFKOKALAFASRGNHVAPTHYVESDA 228  
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db     1914 QMMNRLIAFASRGNHVSPTHYVPESDA 1940

RESULT 11  
US-10-044-995-23  
Sequence 23, Application US/10044995  
Publication No. US20030049685A1  
GENERAL INFORMATION:  
APPLICANT: DELEYS, ROBERT J  
POLLET, DIRK  
MAERTENS, GEERT  
VAN HEUVERSWUD, HUGO  
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF  
ANTIBODIES TO HEPATITIS C VIRUS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAVE P.C.  
STREET: 1100 NORTH GLEEBO ROAD  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/044,995  
FILING DATE: 15-Jan-2002

```
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match      14.0%: Score 254; DB 15; Length 2894;
Best Local Similarity 31.8%; Pred. No. 2,4e-12;
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY 23 RPAVIPDREVLVYOEFDMEDEAS--HLPIYEOGMQLAEOFKOKALAF---ASRGNHVSPTHY 78
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1691 KPAIIPDREVLVREFDEMECSQHLPIYEOGMQLAEOFKOKALGLLOTASRQAEV----- 1745
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSPAHYVPESDASQAAPYIEQ---- 133
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1746 ----IAPAV-----QTMWQKLETFWAGHMNFIISGIQYLAGLSTLPENPAIASIMAF 1793
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 134 -AOVIAHQFKEKYLAF-----ASRGNHDSPTHYVPESDARPAI----- 170
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1794 TAAVTSPTTSQTLNFIILGQWVAQAALAPGAATAFVAGLAGAGVGLGKVLIDILA 1853
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 171 -----IPDREVLVREFDEMEESQHL--PYIEOQMWL-----AE 201
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1854 GYGAGVAGALVAFKIMSGEVPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGEGAV 1913
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 202 QFKOKALAFASRGNHVAPTHYVYESDA 228
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1914 QMNRRLIAFASRGNHVSPTHYVYESDA 1940
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 12
US-09-916-359-2
Sequence 2, Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TREATING C HEPATITIS
FILE REFERENCE: PWC97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
PRIOR FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/388,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 3011
```

```
TYPE: PRT
ORGANISM: Virus
US-09-916-359-2

Query Match      14.0%: Score 254; DB 9; Length 3011;
Best Local Similarity 31.8%; Pred. No. 2,4e-12;
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY 23 RPAVIPDREVLVYOEFDMEDEAS--HLPIYEOGMQLAEOFKOKALAF---ASRGNHVSPTHY 78
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1691 KPAIIPDREVLVREFDEMECSQHLPIYEOGMQLAEOFKOKALGLLOTASRQAEV----- 1745
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSPAHYVPESDASQAAPYIEQ---- 133
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1746 ----IAPAV-----QTMWQKLETFWAGHMNFIISGIQYLAGLSTLPENPAIASIMAF 1793
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 134 -AOVIAHQFKEKYLAF-----ASRGNHDSPTHYVPESDARPAI----- 170
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1794 TAAVTSPTTSQTLNFIILGQWVAQAALAPGAATAFVAGLAGAGVGLGKVLIDILA 1853
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 171 -----IPDREVLVREFDEMEESQHL--PYIEOQMWL-----AE 201
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1854 GYGAGVAGALVAFKIMSGEVPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGEGAV 1913
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 202 QFKOKALAFASRGNHVAPTHYVYESDA 228
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1914 QMNRRLIAFASRGNHVSPTHYVYESDA 1940
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
US-10-259-275-42
Sequence 42, Application US/10259275
Publication No. US20030125541A1
GENERAL INFORMATION:
APPLICANT: Lemon, Stanley M.
TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
FILE REFERENCE: 265,0007 0120
CURRENT APPLICATION NUMBER: US/10/259,275
CURRENT FILING DATE: 2003-01-13
PRIOR APPLICATION NUMBER: US 60/171,909
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: US 09/747,419
PRIOR FILING DATE: 2000-12-23
PRIOR APPLICATION NUMBER: US 60/325,236
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: US 60/338,123
PRIOR FILING DATE: 2001-11-13
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn version 3.0
SEQ ID NO 42
LENGTH: 1985
TYPE: PRT
ORGANISM: ARTIFICIAL
FEATURE:
OTHER INFORMATION: amino acid sequence encoded by the nucleotides 2119-8073 of
US-10-259-275-42
```

```
Query Match      13.8%: Score 251; DB 15; Length 1985;
Best Local Similarity 31.0%; Pred. No. 2,6e-12;
Matches 83; Conservative 30; Mismatches 75; Indels 80; Gaps 10;

QY 23 RPAVIPDREVLVYOEFDMEDEAS--HLPIYEOGMQLAEOFKOKALAF---ASRGNHVSPTHY 78
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 666 KPAIIPDREVLVREFDEMECSQHLPIYEOGMQLAEOFKOKALGLLOTATQAEV----- 720
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSPAHYVPESDASQAAPYIEQ--- 133
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 721 ----AAVVESKRTL-----EAFWAGHMNFIISGIQYLAGLSTLPENPAIASLMA 767
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 134 -AOVIAHQFKEKYLAF-----ASRGNHDSPTHYVPESDARPAI----- 170
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 768 FTASITSPITTOHTLLFNILGWAQAPPSAASFVAGIAGAAGVSGIGUKVLDIL 827  
Qy 171 -----IPDREVLREPFDEMEBSOHL--PYIQGMML-----A 200  
Db 828 AGYGAGVAGALVAFKVMSEMPSTEDLVLLPAITSPGALVGVCAAILRRHVGEGCA 887  
Qy 201 EQFKOKALAFASRGHVAPTHVYTESDA 228  
Db 888 VQMMNRLLIAFASRGHVSPTHVYTESDA 915

RESULT 14  
US-10-232-643-6  
Sequence 6, Application US/10232643  
Publication No. US20030129586A1  
GENERAL INFORMATION:  
APPLICANT: HOUGHTON, MICHAEL  
CHOO, QUI-LIM  
HAN, JANG  
CHOE, JOONHO  
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING  
HELICASE ACTIVITY AND IMPROVED SOLUBILITY  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: Intellectual Property - R440, P.O. Box 8097  
CITY: Emeryville  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/232,643  
FILING DATE: 30-Aug-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/483,799  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/529,169  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Harbin, Alisa A.  
REGISTRATION NUMBER: 33,895  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 923-3274  
TELEFAX: (510) 655-3542  
TELEX: N/A  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3011 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 9  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Lys or Arg"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 11  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Asn or Thr"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 176  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Ile or Thr"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 334  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Met or Val"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 603  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Leu or Ile"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 848  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Tyr or Asn"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 1114  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Pro or Ser"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 1117  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Ser or Thr"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 1276  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Pro or Leu"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 1454  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Cys or Tyr"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 1471  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Thr or Ser"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 1877  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Glu or Gly"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 1948  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Leu or His"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 1949  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Ser or Cys"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 2021  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Gly or Val"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 2349  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Thr or Ser"  
FEATURE:  
NAME/KEY: Duplication  
LOCATION: 2385  
OTHER INFORMATION: /note= "There exists a  
heterogeneity at this position - Xaa = Tyr or Phe"  
FEATURE:  
NAME/KEY: Duplication

RESULT 15  
US-09-238-076-20  
Sequence 20, Application US/099238076  
Patent No. US20020102540A1  
GENERAL INFORMATION:  
APPLICANT: RICE, CHARLES et al.  
TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C  
TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HOWELL & HAFERRAMP, L.C.  
STREET: 7733 FORSYTH BLVD., SUITE 1400  
CITY: ST. LOUIS  
STATE: MO  
COUNTRY: USA  
ZIP: 63105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

Query Match	13.8%	Score 250;	DB 10;	Length 3011;
Best Local Similarity	31.8%	Pred. No. 5.66-12;		
Matches	85;	Conservative	26;	Mismatches 78; Indels 78; Gaps 10;

  

QY	23	RPAVIRPREVLVYQGFEDMEEAAS-HLPRTQSGMQLAEQFKQALAF--ASQGNHVSPTHY	78
Db	1691	KPAIIPREVLVYQGFEDMEESQHLPRTEQGMMLAEQFKQALALLOTASHNAEV----	1745
QY	79	VPESDARPAIIPREVLRHREFDEMEEAFAASR-GNNVSRANHYVPESDASQAARYIEO----	133
Db	1746	-----IIRPAV-----QTNMQKLEVFMAKNNMFISGIQYLAQLSTLRGNRAIASLMAF	1793
QY	134	-AQVIAQRFREKVLAF-----ASQGNHDSPTHYVPESDAKAL-----	170
Db	1794	TAAVTSPLTTGGQTLLENILGSGWAALDARPAALTAFAAGAGLAGAAGISVGLKVLVDILA	1853
QY	171	-----IPREVLVREFDEMEESQHL--PYIEQGMML-----AE	201
Db	1854	GYGAGVAGALVAFKIMSGEVPSTEDLVNLLRALISPGALVVGVCVCAIIRRNVRGEGAV	1913
QY	202	QFKQKALAFASRGNHVAPTHYVPESDA	228
Db	1914	QMMNRLIAFASRGNHVSPTHYVPESDA	1940

Search completed: July 25, 2003, 13:21:07  
Job time : 54 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 13:08:21 ; Search time 30 Seconds

(without alignments)  
496.447 Million cell updates/sec

Title: US-09-491-146a-52

Perfect score: 1818  
Sequence: 1 AAHIPYLEQGMHLEQFKOK.....KALIEEGQMAEMLSKIQ 352

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

a1 number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1818	100.0	352	US-08-921-887-52	Sequence 52, Appli
2	257	14.1	3011	US-08-188-281B-1	Sequence 1, Appli
3	257	14.1	3011	US-08-453-552-1	Sequence 1, Appli
4	257	14.1	3011	US-08-710-637-1	Sequence 1, Appli
5	257	14.1	3011	PCT-US93-00907-1	Sequence 1, Appli
6	257	14.1	3011	PCT-US94-07280-1	Sequence 1, Appli
7	257	14.1	3011	PCT-US95-01087-1	Sequence 1, Appli
8	256	14.1	2013	US-08-324-977-12	Sequence 12, Appli
9	256	14.1	2013	US-08-384-616-12	Sequence 12, Appli
10	256	14.1	2013	US-08-904-686A-12	Sequence 12, Appli
11	256	14.1	2013	US-09-315-850-12	Sequence 12, Appli
12	256	14.1	2201	US-08-952-981A-2	Sequence 2, Appli
13	256	14.1	2620	US-08-324-977-32	Sequence 32, Appli
14	256	14.1	2620	US-08-384-616-32	Sequence 32, Appli
15	256	14.1	2620	US-08-904-686A-32	Sequence 32, Appli
16	256	14.1	2620	US-09-315-850-32	Sequence 32, Appli
17	256	14.1	2621	US-08-324-977-36	Sequence 36, Appli
18	256	14.1	2621	US-08-384-616-36	Sequence 36, Appli
19	256	14.1	2621	US-08-904-686A-36	Sequence 36, Appli
20	256	14.1	2621	US-09-315-850-36	Sequence 36, Appli
21	256	14.1	3010	US-08-324-977-2	Sequence 2, Appli
22	256	14.1	3010	US-08-324-977-14	Sequence 14, Appli
23	256	14.1	3010	US-08-384-616-2	Sequence 2, Appli
24	256	14.1	3010	US-08-384-616-14	Sequence 14, Appli
25	256	14.1	3010	US-08-904-686A-2	Sequence 2, Appli
26	256	14.1	3010	US-08-904-686A-14	Sequence 14, Appli
27	256	14.1	3010	US-09-315-850-2	Sequence 2, Appli

28	256	14.1	3010	US-09-315-850-14	Sequence 14, Appli
29	255	14.0	3012	US-08-811-566-2	Sequence 2, Appli
30	255	14.0	3012	US-09-034-756-2	Sequence 2, Appli
31	254	14.0	360	US-08-850-328-4	Sequence 4, Appli
32	254	14.0	859	US-08-444-818-30	Sequence 30, Appli
33	254	14.0	2261	US-08-444-818-66	Sequence 66, Appli
34	254	14.0	2772	US-08-444-818-89	Sequence 89, Appli
35	254	14.0	2894	US-08-466-975A-23	Sequence 23, Appli
36	254	14.0	2894	US-08-391-671A-23	Sequence 23, Appli
37	254	14.0	2894	US-08-467-902A-23	Sequence 23, Appli
38	254	14.0	2894	US-09-275-265-23	Sequence 23, Appli
39	254	14.0	2894	US-09-941-611-23	Sequence 23, Appli
40	254	14.0	2895	US-08-444-818-138	Sequence 138, Appli
41	254	14.0	3011	US-08-440-103-36	Sequence 36, Appli
42	254	14.0	3011	US-08-440-542-36	Sequence 36, Appli
43	254	14.0	3011	US-07-910-760-10	Sequence 10, Appli
44	254	14.0	3011	US-08-440-519-10	Sequence 10, Appli
45	254	14.0	3011	US-08-231-368-36	Sequence 36, Appli

## ALIGNMENTS

RESULT 1  
US-08-921-887-52  
Sequence 52, Application US/08921887  
Parent No. 6030771  
GENERAL INFORMATION:  
APPLICANT: KHUDYAKOV, YURI E.  
APPLICANT: FIELDS, HOWARD A.  
TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION  
TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JONES & ASKEW, LLP  
STREET: 191 Peachtree Street, N.W., 37th Floor  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30303-1769  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/921,887  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: WARREN, WILLIAM L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 03063-0380  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 352 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: Protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: Hepatitis virus  
US-08-921-887-52

Query Match 100.0%; Score 1818; DB 3; Length 352;  
Best Local Similarity 100.0%; Pred. No. 7,6e-163;  
Matches 352; Conservative 0; Mismatches 0; Gaps 0;

QY 1 AAHPIYLEOGMHLAEQFKOKALRPVTPDREVLVOEFDEMEASHLPIYEOGMQLAEQFK 60  
DB 1 AAHPIYLEOGMHLAEQFKOKALRPVTPDREVLVOEFDEMEASHLPIYEOGMQLAEQFK 60  
QY 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLHREFDEMEAFASRGNHVSPTHYVE 120  
DB 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLHREFDEMEAFASRGNHVSPTHYVE 120  
QY 121 ESASQAAPYIEQAQVIAHOFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLHRE 180  
DB 121 ESASQAAPYIEQAQVIAHOFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLHRE 180  
QY 181 FDEMEASHLPIYEOGMHLAEQFKOKALAFASRGNHVAPTHYTESDAKALVPDKEVLY 240  
DB 181 FDEMEASHLPIYEOGMHLAEQFKOKALAFASRGNHVAPTHYTESDAKALVPDKEVLY 240  
QY 241 QOYDEMEAFASRGNHVAPTHYVESDASASLPMDETALIAQFKEKVLAFASRGNHVS 300  
DB 241 QOYDEMEAFASRGNHVAPTHYVESDASASLPMDETALIAQFKEKVLAFASRGNHVS 300  
QY 301 PRHYVESEPOVVVTPDKELVFAFDEMEASKALIEGQMAEWLKSQIQ 352  
DB 301 PRHYVESEPOVVVTPDKELVFAFDEMEASKALIEGQMAEWLKSQIQ 352

## RESULT 2

US-08-188-281B-1  
Sequence 1, Application US/08188281B  
Patent No. 5610009  
GENERAL INFORMATION:  
APPLICANT: MATANABE, SHINICHI  
APPLICANT: YAMAGUCHI, JULIE  
APPLICANT: DESAI, SURESH M.  
APPLICANT: DEVARE, SUSHIL G.  
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV  
TITLE OF INVENTION: ENVELOPE GENES  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: ONE ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/188,281B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FOREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5521.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-937-6365  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3011 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-188-281B-1

Query Match 14.1%, Score 257, DB 1, Length 3011,  
Best Local Similarity 32.5%, Pred. No. 7,4e-15;  
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVTPDREVLVOEFDEMEAS-HLPIYEOGMQLAEQFKOKALAF---ASRGNHVSPTHY 78  
DB 1691 KPAIIPDREVLVOEFDEMEASQHLPIYEOGMHLAEQFKOKALGLQTAASQAQV----- 1745  
QY 79 VESDARPAIIPDREVLHREFDEMEAFASR---GNHVSPTHYVESDASQAAPYIEQ--- 133  
DB 1746 -----ITPAVQTNMOKL-----EATWAKMNNFISGTOYLAGLSTLPGNPAIASLMA 1792  
QY 134 ---AQVIAHOFKEKVLAF-----ASRGNHVSPTHYVESDARPAI-----A 170  
DB 1793 FTAATVSTLTSQTLTLFNILGQVAAQLAARPAIATFAGAGLAGAIISSVGLKVLVDIL 1852  
QY 171 -----IPDREVLHREFDEMEASQHL---PIYEOGMHL-----A 200  
DB 1853 AGYGAGVALVAFKIMSGEVPSTEDLVNLLPALISPALVGVCAAILRRHVGPEGCA 1912  
QY 201 EGFKOKALAFASRGNHVAPTHYVESDA 228  
DB 1913 VQMMNRLIAFASRGNHVSPTHYVESDA 1940

## RESULT 3

US-08-453-552-1  
Sequence 1, Application US/08453552  
Patent No. 5667992  
GENERAL INFORMATION:  
APPLICANT: CASEY, JAMES M.  
APPLICANT: BODE, SUZANNE L.  
APPLICANT: ZECK, BILLY J.  
APPLICANT: YAMAGUCHI, JULIE  
APPLICANT: FRAIL, DONALD E.  
APPLICANT: DESAI, SURESH M.  
APPLICANT: DEVARE, SUSHIL G.  
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV  
TITLE OF INVENTION: PROTEINS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: ONE ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,552  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FOREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5131.US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9556  
TELEFAX: 708-937-9556  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3011 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-552-1

Query Match 14.1%, Score 257, DB 1, Length 3011,  
Best Local Similarity 32.5%, Pred. No. 7,4e-15;  
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;



QY 23 RPVLPDREVLYQGFDEMEEAS-HLPYIEQGMQLAEQFKOKALAF---ASRGNHVSPTHY 78  
Db 1691 KPAIIPDREVLYQGFDEMEECQHLPYIEQGMQLAEQFKOKALGLQTAASQAQAEV----- 1745  
QY 79 VPESDARPAIIPDREVLYHREFDEMEEAFASR--GNHVSAPHYVPESDASQAAPYIEQ--- 133  
Db 1746 -----ITPAVQTNMOKL-----EAFWAKHMMNFISGTOYLAGLSTLPGNPAIASLMA 1792  
QY 134 --AOVIAHQFEKVLAF-----ASRGNHDSPTHYVPESDAPAI----- 170  
Db 1793 FTAAVTSPLTTSQTLFNLIGGWAAQLAARGAATATFAGAGLAGAAGISVGLGKVLVDIL 1852  
QY 171 -----IPDREVLYREFDEMEESQHL--PYIEQGMQL-----A 200  
Db 1853 AGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAIISPGLALVGVCAAILRRHVGPGEGA 1912  
QY 201 EQFKOKALAFASRGHVAPTHYVTESDA 228  
Db 1913 VQMMNRLLAFASRGHVSPTHYVPESDA 1940

-----JUL 4  
US-08-710-637-1  
Sequence 1, Application US/08710637  
Patent No. 5854001

GENERAL INFORMATION:  
APPLICANT: CASEY, JAMES M.  
APPLICANT: BODE, SUZANNE L.  
APPLICANT: ZECK, BILLY J.  
APPLICANT: YAMAGUCHI, JULIE  
APPLICANT: FRAIL, DONALD E.  
APPLICANT: DESAI, SURESH M.  
APPLICANT: DEVARE, SUSHIL G.  
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV  
NUMBER OF INVENTION: 12  
NUMBER OF SEQUENCES: 12  
TITLE OF INVENTION: PROTEINS  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: ONE ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/710,637  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/144,099  
FILING DATE:  
APPLICATION NUMBER: US 07/830,024  
FILING DATE: 01-JAN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: FOREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5111.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-937-9556  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3011 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-710-637-1

Query Match 14.1%; Score 257; DB 2; Length 3011;  
Best Local Similarity 32.5%; Pred. No. 7.4e-15;  
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;  
QY 23 RPVLPDREVLYQGFDEMEEAS-HLPYIEQGMQLAEQFKOKALAF---ASRGNHVSPTHY 78  
Db 1691 KPAIIPDREVLYQGFDEMEECQHLPYIEQGMQLAEQFKOKALGLQTAASQAQAEV----- 1745  
QY 79 VPESDARPAIIPDREVLYHREFDEMEEAFASR--GNHVSAPHYVPESDASQAAPYIEQ--- 133  
Db 1746 -----ITPAVQTNMOKL-----EAFWAKHMMNFISGTOYLAGLSTLPGNPAIASLMA 1792  
QY 134 --AOVIAHQFEKVLAF-----ASRGNHDSPTHYVPESDAPAI----- 170  
Db 1793 FTAAVTSPLTTSQTLFNLIGGWAAQLAARGAATATFAGAGLAGAAGISVGLGKVLVDIL 1852  
QY 171 -----IPDREVLYREFDEMEESQHL--PYIEQGMQL-----A 200  
Db 1853 AGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAIISPGLALVGVCAAILRRHVGPGEGA 1912  
QY 201 EQFKOKALAFASRGHVAPTHYVTESDA 228  
Db 1913 VQMMNRLLAFASRGHVSPTHYVPESDA 1940

-----RESULT 5  
PCT-US93-00907-1  
Sequence 1, Application PC/TUS9300907

GENERAL INFORMATION:  
APPLICANT: CASEY, JAMES M.  
APPLICANT: BODE, SUZANNE L.  
APPLICANT: ZECK, BILLY J.  
APPLICANT: YAMAGUCHI, JULIE  
APPLICANT: FRAIL, DONALD E.  
APPLICANT: DESAI, SURESH M.  
APPLICANT: DEVARE, SUSHIL G.  
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV  
NUMBER OF INVENTION: 12  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: ONE ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/00907  
FILING DATE: 19930129  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: FOREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5131.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9556  
TELEFAX: 708-937-6365  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3011 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-00907-1

Query Match 14.1%; Score 257; DB 5; Length 3011;  
Best Local Similarity 32.5%; Pred. No. 7.4e-15;

Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVIPDREVLYOEFDEMEAS-HLPIYEQMOLAEOFKOKALAF--ASRONHVSPTHY 78  
Db 1691 KPAIIPDREVLYOEFDEMECSOHLPIYEQMMLAEQFKOALGLQTSRQAEV----- 1745

QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSAPHYVPESDASQAAPYIEQ--- 133  
Db 1746 -----ITPAVQTNMOKL-----EAFWAKHMNFIISGTQYLAGLSTLPGNPAIASLMA 1792

QY 134 --AOVIAHQFEKVLAF-----ASRGNDSPTHYVPESDAKKPAI----- 170  
Db 1793 FTAAYVSPLTTSQTLFNLIGGVAAQLAPGAATAFVAGLAGAIGSVGLKVLVDIL 1852

QY 171 -----IPDREVLYREFDEMECSOHL--PYIEQGMML-----A 200  
Db 1853 AGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAIISPGLVGVVCAAILRRHVGPGEA 1912

QY 201 EQFKOKALAFASRGNHVAPTHYTESDA 228  
1913 VQMMNRLIAFASRGNHVSPTHYVPESDA 1940

RESULT 6  
PCT-US94-07280-1  
; Sequence 1, Application PC/TUS9407280  
; GENERAL INFORMATION:  
; APPLICANT: YAMAGUCHI, SHINICHI  
; APPLICANT: YAMAGUCHI, JULIE  
; APPLICANT: DESAI, SURESH M.  
; APPLICANT: DEVARE, SUSHIL G.  
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
; STREET: ONE ABBOTT PARK ROAD  
; CITY: ABBOTT PARK  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/07280  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: POREMSKI, PRISCILLA E.  
; REGISTRATION NUMBER: 33,207  
; REFERENCE/DOCKET NUMBER: 5521.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 708-937-6365  
; TELEFAX: 708-938-2623  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3011 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US94-07280-1

Query Match 14.1%; Score 257; DB 5; Length 3011;  
Best Local Similarity 32.5%; Pred. No. 7,4e-15;  
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVIPDREVLYOEFDEMEAS-HLPIYEQMOLAEOFKOKALAF--ASRONHVSPTHY 78  
Db 1691 KPAIIPDREVLYOEFDEMECSOHLPIYEQMMLAEQFKOALGLQTSRQAEV----- 1745

QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSAPHYVPESDASQAAPYIEQ--- 133  
Db 1746 -----ITPAVQTNMOKL-----EAFWAKHMNFIISGTQYLAGLSTLPGNPAIASLMA 1792

QY 134 --AOVIAHQFEKVLAF-----ASRGNDSPTHYVPESDAKKPAI----- 170  
Db 1793 FTAAYVSPLTTSQTLFNLIGGVAAQLAPGAATAFVAGLAGAIGSVGLKVLVDIL 1852

QY 171 -----IPDREVLYREFDEMECSOHL--PYIEQGMML-----A 200  
Db 1853 AGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAIISPGLVGVVCAAILRRHVGPGEA 1912

QY 201 EQFKOKALAFASRGNHVAPTHYTESDA 228  
1913 VQMMNRLIAFASRGNHVSPTHYVPESDA 1940

RESULT 7  
PCT-US95-01087-1  
; Sequence 1, Application PC/TUS9501087  
; GENERAL INFORMATION:  
; APPLICANT: YAMAGUCHI, SHINICHI  
; APPLICANT: YAMAGUCHI, JULIE  
; APPLICANT: DESAI, SURESH M.  
; APPLICANT: DEVARE, SUSHIL G.  
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
; STREET: ONE ABBOTT PARK ROAD  
; CITY: ABBOTT PARK  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/01087  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: POREMSKI, PRISCILLA E.  
; REGISTRATION NUMBER: 33,207  
; REFERENCE/DOCKET NUMBER: 5521.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 708-937-6365  
; TELEFAX: 708-938-2623  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3011 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-01087-1

Query Match 14.1%; Score 257; DB 5; Length 3011;  
Best Local Similarity 32.5%; Pred. No. 7,4e-15;  
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVIPDREVLYOEFDEMEAS-HLPIYEQMOLAEOFKOKALAF--ASRONHVSPTHY 78  
Db 1691 KPAIIPDREVLYOEFDEMECSOHLPIYEQMMLAEQFKOALGLQTSRQAEV----- 1745

QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSAPHYVPESDASQAAPYIEQ--- 133  
Db 1746 -----ITPAVQTNMOKL-----EAFWAKHMNFIISGTQYLAGLSTLPGNPAIASLMA 1792

QY 134 --AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVESPDAKPAI----- 170  
 DB 1793 FTAATVSTPLTTSQTLFFNLGCVAAQLAPGAATAFVAGAGLAGAAGISVGKLVLDL 1852  
 QY 171 -----IPREVLYREFDEMEESQHL--PYIEQGMML-----A 200  
 DB 1853 AGYAGVAGALVAFKVMGSEVPSTEDLVNLLPALISPGALVGVVCAAILRRHVGPGECA 1912  
 QY 201 EGFKOKALAFASRGNHVAPTHYVETSDA 228  
 DB 1913 VQMNRLIFASRGNHVSPTHYVESPDA 1940

RESULT 8  
 US-08-324-977-12  
 ; Sequence 12, Application US/08324977  
 ; Patent No. 5747339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OKAYAMA, Hiroto  
 ; APPLICANT: FUKE, Isao  
 ; APPLICANT: MORI, Chisato  
 ; APPLICANT: TAKAMIZAWA, Akahisa  
 ; APPLICANT: YOSHIDA, Iwao  
 ; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC  
 ; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE  
 ; NUMBER OF SEQUENCES: 50  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Armstrong, Westerman, Hattori, Mclelland &  
 ; STREET: 1725 K St. N.W. Suite 1000  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20006  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/324,977  
 ; FILING DATE: 18-OCT-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 2-167466  
 ; FILING DATE: 25-JUN-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 2-230921  
 ; FILING DATE: 31-AUG-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 2-305605  
 ; FILING DATE: 09-NOV-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/099,706  
 ; FILING DATE: 30-JUL-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/769,996  
 ; FILING DATE: 02-OCT-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/635,451  
 ; FILING DATE: 28-DEC-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Stevens-Smith, Theresa M.  
 ; REGISTRATION NUMBER: 36,281  
 ; REFERENCE/DOCKET NUMBER: 900703D  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 659-2930  
 ; TELEFAX: (202) 887-0357  
 ; TELEX: 440142  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2013 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear

MOLECULE TYPE: protein  
 US-08-324-977-12  
 Query Match 14.1%; Score 256; DB 1; Length 2013;  
 Best Local Similarity 31.1%; Pred. No. 5.1e-15;  
 Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPAVDPREVLYOEEDMEEB-ASHLPIYEQGMQLAEQKOKALAF---ASRGNHVSPTHY 78  
 DB 1691 RPAVDPREVLYOEEDMEEBQSHLPIYEQGMQLAEQKOKALAGLLOTAIKOAE----- 1745  
 QY 79 VPESDARPAIIPREVLYREFDEMEEAFASR-GNVSPAHYVESPDAQOAPYIEQ----- 133  
 DB 1746 -----AAP-----VVEASKRALETFAKGMNMFISGIQLAGLSTLPGPALASLMAF 1793  
 QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVESPDAKPAI----- 170  
 DB 1794 TASITSPLTTSQTLFFNLGCVAAQLAPPSAASFVAGAGIAGAVGSGIGKLVLDILA 1853  
 QY 171 -----IPREVLYREFDEMEESQHL--PYIEQGMML-----AE 201  
 DB 1854 GYAGVAGALVAFKVMGSEVPSTEDLVNLLPALISPGALVGVVCAAILRRHVGPGEAV 1913  
 QY 202 QFKOKALAFASRGNHVAPTHYVETSDA 228  
 DB 1914 QMNRLIFASRGNHVSPTHYVESPDA 1940

RESULT 9  
 US-08-384-616-12  
 ; Sequence 12, Application US/08384616  
 ; Patent No. 5847101  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OKAYAMA, Hiroto  
 ; APPLICANT: FUKE, Isao  
 ; APPLICANT: MORI, Chisato  
 ; APPLICANT: TAKAMIZAWA, Akahisa  
 ; APPLICANT: YOSHIDA, Iwao  
 ; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC  
 ; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE  
 ; NUMBER OF SEQUENCES: 50  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Armstrong, Westerman, Hattori, Mclelland &  
 ; STREET: 1725 K St. N.W. Suite 1000  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20006  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/384,616  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/769,996  
 ; FILING DATE: 02-OCT-1991  
 ; APPLICATION NUMBER: JP 2-167466  
 ; FILING DATE: 25-JUN-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 2-230921  
 ; FILING DATE: 31-AUG-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 2-305605  
 ; FILING DATE: 09-NOV-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/635,451  
 ; FILING DATE: 28-DEC-1990  
 ; ATTORNEY/AGENT INFORMATION:









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 25, 2003, 12:54:06 ; Search time 85 Seconds

(without alignments)  
657.315 Million cell updates/sec

Title: US-09-491-146a-52

Perfect score: 1818  
Sequence: 1 AAHPYLEQGMHLAEQFKQ.....KALIEEGQMAEWLKSIIQ 352

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

a1 number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1818	100.0	352	20	AAV06672
2	512.5	28.2	829	23	AAE18690
3	512.5	28.2	1099	23	AAU76378
4	512.5	28.2	1099	24	ABG72262
5	268	14.7	3010	14	AAK30616
6	263	14.5	3010	16	AAK2694
7	263	14.5	3011	14	AAK34468
8	260	14.3	3010	16	AAK6864
9	259	14.2	1786	10	AAK90158

10	259	14.2	2354	14	AAK41435
11	259	14.2	2510	13	AAK29527
12	259	14.2	3011	15	AAK66995
13	258	14.2	1250	12	AAK12599
14	257	14.1	3010	15	AAK3417
15	257	14.1	3011	16	AAK6622
16	257	14.1	3011	16	AAK40119
17	257	14.1	3011	16	AAK79232
18	256	14.1	2201	16	AAK01680
19	256	14.1	3010	13	AAK20091
20	256	14.1	3010	13	AAK20111
21	256	14.1	3010	20	AAV06423
22	256	14.1	3011	16	AAK67588
23	255	14.0	1411	13	AAK29533
24	255	14.0	3011	24	AAK77397
25	255	14.0	3011	24	ABP71460
26	255	14.0	3012	23	AAU99289
27	254	14.0	858	10	AAK90146
28	254	14.0	859	10	AAK92029
29	254	14.0	1766	10	AAK92041
30	254	14.0	1771	22	AAK62631
31	254	14.0	1771	22	AAK62632
32	254	14.0	1771	22	AAK62634
33	254	14.0	1771	22	AAK62635
34	254	14.0	1892	22	AAK62636
35	254	14.0	1911	22	AAK62638
36	254	14.0	1921	22	AAK62639
37	254	14.0	1944	22	AAK62637
38	254	14.0	2201	23	ABG30583
39	254	14.0	2201	23	ABG30586
40	254	14.0	2201	23	ABG30588
41	254	14.0	2261	10	AAK90164
42	254	14.0	2301	10	AAK92047
43	254	14.0	2435	13	AAK25135
44	254	14.0	2436	10	AAK92050
45	254	14.0	2436	10	AAK92088

#### ALIGNMENTS

RESULT 1  
AAV06672 standard; Protein; 352 AA.  
AC AAV06672;  
XX 17-JUN-1999 (first entry)  
DT  
XX Amino acid sequence of the NS4 mosaic protein.  
XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REA;  
KW restriction endonuclease assisted ligation; vaccination; NS4 antigen.  
XX  
XX Hepatitis C virus.  
OS Synthetic.  
XX  
XX WO9910506-A1.  
XX  
XX 04-MAR-1999.  
XX  
XX 21-AUG-1998; 98WO-US17385.  
XX 25-AUG-1997; 97US-0921887.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX Fields HA, Khudiyakov YE;  
XX WPI; 1999-204671/17.  
XX  
XX New mosaic protein, comprising a plurality of homologous antigenic  
PT peptides from different genotypes of a species - useful for

PT-NANBH virus non  
HCV antigen T7N1-3  
Hepatitis C virus  
Portion of PT-NANBH  
Blood transmissible  
HCV protein cleava  
HCV genomic amino  
HCV sequence. Hep  
HCV NS2-NS5B non-S  
Non-A, non-B viral  
Non-A, non-B viral  
Non-A, non-B hepat  
Hepatitis C virus  
HCV NS4-NS5 peptid  
Hepatitis C virus  
Amino acid sequenc  
Hepatitis C virus  
ORF extending thro  
Sequence encoded i  
Sequence encoded i  
HCV NS3s polypepti  
HCV deINS3s polype  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Hepatitis C virus  
Hepatitis C virus  
Hepatitis C virus  
Peptide encoded by  
Sequence encoded i  
HCV polypeptide 1.  
Sequence encoded i  
Peptide encoded by



PT detecting hepatitis infection in an individual  
XX  
XX Example 3; Fig 18; 66pp; English.  
PS  
XX The invention relates to a mosaic protein, comprising a plurality of  
CC homologous antigenic peptides from different genotypes of a species. The  
CC antigenic peptides are from nucleocapsid (NC) proteins. A method for  
CC synthesizing an artificial gene that encodes the mosaic protein is also  
CC provided. The method is designated restriction endonuclease assisted  
CC ligation (REAL). The mosaic protein and the artificial mosaic protein  
CC are useful for detecting a hepatitis infection in an individual. The  
CC mosaic gene and protein is also useful for vaccination against infection,  
CC especially hepatitis C. The method of synthesizing the artificial gene  
CC and the resulting mosaic protein improve the sensitivity, spectrum of  
CC immunoreactivity, and antigen specificity of enzyme immunoassays. This  
CC provides improved detection of hepatitis C virus. The present sequence  
CC represents the amino acid sequence of the NS4 mosaic protein.  
XX  
SO Sequence 352 AA:  
  
Query Match 100.0%; Score 1818; DB 20; Length 352;  
Best Local Similarity 100.0%; Pred. No. 1.8e-148;  
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAHPIPLEQGMHLAEQFKOKALRPAYIPREVLYQGFDEMEASHLPYIEQGMQLAEQPK 60  
DB 1 AAHPIPLEQGMHLAEQFKOKALRPAYIPREVLYQGFDEMEASHLPYIEQGMQLAEQPK 60  
QY 61 OKALAFASRGNHVSPTHYVPESDARPAIIPREVLYHREFDEMEAFASRGNHVSFAHYVP 120  
DB 61 OKALAFASRGNHVSPTHYVPESDARPAIIPREVLYHREFDEMEAFASRGNHVSFAHYVP 120  
QY 121 ESDASQAAPYIEQAQVIAHQFKEKVLAFASRGNHDSPTHYVPESDAKPAIIPREVLYRE 180  
DB 121 ESDASQAAPYIEQAQVIAHQFKEKVLAFASRGNHDSPTHYVPESDAKPAIIPREVLYRE 180  
QY 181 FEMESQHLPIYIEQGMHLAEQFKOKALAFASRGNHVAPTHYVTSDDAPALVPKEXVY 240  
DB 181 FEMESQHLPIYIEQGMHLAEQFKOKALAFASRGNHVAPTHYVTSDDAPALVPKEXVY 240  
QY 241 QOYDEMEAFASRGNHVAPTHYVTSDDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300  
DB 241 QOYDEMEAFASRGNHVAPTHYVTSDDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300  
QY 301 PPHYVPESRPQVVVTPDKELIYEAFDEMEASKALIEGQMAEMLSKIQ 352  
DB 301 PPHYVPESRPQVVVTPDKELIYEAFDEMEASKALIEGQMAEMLSKIQ 352  
  
-SULT 2  
AAE18690  
ID AAE18690 standard; Protein; 829 AA.  
XX  
AC AAE18690;  
XX  
XX 17-MAY-2002 (first entry)  
DT  
XX  
DE Multiple epitope fusion antigen (MEFA) 12 protein.  
XX  
XX Hepatitis C virus; NS3/4a antigen; multiple epitope fusion antigen;  
KW HCV infection; MEFA 12 protein.  
XX  
OS Unidentified.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 315 /note= "Encoded by ATG"  
FT Misc-difference 645 /note= "Encoded by GAG"  
TT  
XX  
XX WO200196875-A2.  
XX  
XX PD 20-DEC-2001.

XX  
PF 14-JUN-2001; 2001WO-US19369.  
XX  
XX 15-JUN-2000; 2000US-212082P.  
PR 02-APR-2001; 2001US-280811P.  
PR 02-APR-2001; 2001US-280867P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Chien DY, Arcangel P, Tandeske L, George-Nascllemento C, Colt D;  
PI Medina-Selby A;  
PI  
XX  
DR MPI: 2002-179522/23.  
DR N-PSDB: AAD29796.  
XX  
XX Immunoassay solid support useful for detecting hepatitis C virus  
PT infection in a biological sample, comprises at least one of HCV  
PT anti-core antibody and HCV NS3/4a epitope, bound to the support  
XX  
PS Disclosure: Fig 7; 87pp; English.  
XX  
XX The present invention relates to hepatitis C virus (HCV) core antigen  
CC and NS (nonstructural) 3/4a antibody combination assay that can detect  
CC both HCV antigens and antibodies present in a sample using a single  
CC solid matrix as well as immunoassay solid supports for use in the assay.  
CC The solid support is useful for detecting HCV infection in a biological  
CC sample. The present sequence is MEFA (multiple epitope fusion antigen)  
CC 12 protein. This sequence is used in the exemplification of the  
CC invention.  
XX  
SQ Sequence 829 AA:  
  
Query Match 28.2%; Score 512.5; DB 23; Length 829;  
Best Local Similarity 62.8%; Pred. No. 1.9e-35;  
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;  
  
QY 167 KPAILIPREVLYREFDEMEE-SCHLPYIEQGMHLAEQFKOKALAFASRGNHVAPTHYVTE 225  
DB 397 KPAILIPREVLYREFDEMECSQHLPIYIEQGMHLAEQFKOKALGL----- 444  
QY 226 SDAPALVPDKEXVYQOYDEMEAFASRGNHVAPTHYVTSDDASASLPYMDETRAIAGQF 285  
DB 445 --GKPAIVPDKEXVYQOYDEME-----CSQAAPYIEQAQVIAHQF 483  
QY 286 KEKVLAFASRGNHVSPPHYVPESRPQVVVTPDKELIYEAFDEME-ASKALIEGQMA 344  
DB 484 KEKVLGLI-----DNDQVVVTPDKELIYEAFDEMECASKALIEGQMA 529  
QY 345 EMLSKIQ 352  
DB 530 EMLSKIQ 537  
  
RESULT 3  
AAU76378  
ID AAU76378 standard; Protein; 1099 AA.  
XX  
AC AAU76378;  
XX  
XX 08-MAY-2002 (first entry)  
DT  
XX  
DE HCV multiple epitope fusion antigen (MEFA) 7.1 protein sequence.  
XX  
XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;  
KW immunoassay solid support; multiple epitope fusion antigen; MEFA;  
KW non-structural protein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
XX WO200196870-A2.  
XX  
XX PD 20-DEC-2001.

```

XX PF 14-JUN-2001; 2001WO-US19156.
XX PR 15-JUN-2000; 2000US-212082P.
XX PR 02-APR-2001; 2001US-280811P.
XX PR 02-APR-2001; 2001US-280867P.
XX PA (CHIR ) CHIRON CORP.
XX PI Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D,
XX PI Medina-selby A;
XX DR WPI; 2002-090228/12.
XX DR N-PSDB; ABK15345.
XX PT Immunoassay solid support, useful for detecting hepatitis C virus
XX PT infection in biological sample, comprises HCV NS3/4a conformational
XX PT epitope and multiple epitope fusion antigen bound to the support
XX XX
XX Claim 5; Fig 5; 92pp; English.
XX
CC The present invention relates to a new immunoassay solid support
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC conformational epitope and a multiple epitope fusion antigen (MEFA),
CC bound to the support. The NS3/4a conformational epitope and/or
CC MEFA reacts specifically with anti-HCV antibodies present in a biological
CC sample from an HCV-infected individual. The immunoassay of the invention
CC is useful for detecting hepatitis C virus infection in a biological
CC sample. The method of the invention provides a sensitive, accurate
CC diagnostic and prognostic tool to provide adequate patient care and to
CC prevent transmission of HCV by blood and by blood products, or by
CC personal contact. Use of NS3/4a conformational epitope in combination
CC with MEFA, provides a sensitive and reliable method for detecting early
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC masking problems, improving sensitivity in detecting antibodies by
CC allowing a greater number of epitopes on a unit surface area of
CC substrate, and improving substrate. Detection accuracy is increased and
CC the incidence of false results is reduced because of the identification
CC and the use of highly immunogenic HCV antigens which are present during
CC the early stages of HCV seroconversion. The present amino acid sequence
CC represents the multiple epitope fusion antigen (MEFA) 7.1 of the
CC invention.
XX CC
XX SQ Sequence 1099 AA;
XX
Query Match 28.2%; Score 512.5; DB 23; Length 1099;
Best Local Similarity 62.8%; Pred. No. 2,8e-35;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;
167 KPAIIPDREVLVREFDEME--SOHLPTIEQGMMLAEQFOKALAFASRGNHVAPTHVYTE 225
DB 701 KPAIIPDREVLVREFDEMECSOHLPTIEQGMMLAEQFOKALGL----- 748
OY 236 SDAPALVPDEKVLVQYVDEMEAFASRGNHVAPTHVVDASASLPYMETRAIAQGF 285
DB 749 --GKPAIVPDKREVLYQYVDEME-----CSQAAPVIEQAQVIAHQF 787
OY 286 KKKVLAFAFASRGNHVBRHVPESPEQVVTTPDKETILYEAFDEMEF--ASGAALIEGQMA 344
DB 788 KKKVGLI-----DNDQVVVTPDKETILYEAFDEMECASGAALIEGQMA 833
OY 345 EMLKSKIQ 352
DB 834 EMLKSKIQ 841

```

```

XX DE HCV multiple epitope fusion antigen 7.1 (MEFA 7.1).
XX XX
XX XX Immunoassay solid support; Hepatitis C Virus type-1; HCV-1; HCV-2;
XX KM NS3/4a conformational epitope; multiple epitope fusion antigen 7.1;
XX KM MEFA 7.1; anti-HCV antibody; NS3/4a conformational antigen; HCV-3;
XX KM HCV infection; Hepatitis C Virus type-2; Hepatitis C Virus type-3;
XX KM mutant; mutein.
XX XX
XX OS Chimeric - Hepatitis C virus type 1.
XX OS Chimeric - Hepatitis C virus type 2.
XX OS Chimeric - Hepatitis C virus type 3.
XX OS Synthetic.
XX XX
XX Key Location/Qualifiers
XX FH 1..156
XX FT Region /note= "Correspond to amino acids 1-156 of HCV-1 hNSD
XX FT superoxide dismutase"
XX FT
XX FT Region 159..176
XX FT /note= "Correspond to amino acids 303-320 of HCV-1 E1"
XX FT 179..199
XX FT /note= "Correspond to consensus sequence of amino acids
XX FT 390-410 of HCV-1 E2 HVR"
XX FT
XX FT Region 200..230
XX FT /note= "Correspond to consensus sequence of amino acids
XX FT 384-414 of HCV-1 and HCV-2 E2 HVR"
XX FT
XX FT Region 231..696
XX FT /note= "Correspond to amino acids 1193-1658 of HCV-1
XX FT helicase"
XX FT
XX FT Region 699..745
XX FT /note= "Correspond to amino acids 1689-1735 of HCV-1
XX FT 5-1-1 epitope"
XX FT
XX FT Region 748..794
XX FT /note= "Correspond to amino acids 1689-1735 of HCV-3
XX FT 5-1-1 epitope"
XX FT
XX FT Region 797..843
XX FT /note= "Correspond to amino acids 1689-1735 of HCV-2
XX FT 5-1-1 epitope"
XX FT
XX FT Region 846..881
XX FT /note= "Correspond to amino acids 1901-1936 of HCV-1
XX FT polypeptide C100"
XX FT
XX FT Region 884..919
XX FT /note= "Correspond to amino acids 2278-2313 of HCV-1
XX FT NS5 region"
XX FT
XX FT Region 922..957
XX FT /note= "Correspond to amino acids 2278-2313 of HCV-1
XX FT NS5 region"
XX FT
XX FT Region 958..1028
XX FT /note= "Correspond to amino acids 2278-2313 of HCV-1
XX FT NS5 region"
XX FT
XX FT Region 1029..1099
XX FT /note= "Correspond to core region antigenic determinants
XX FT from amino acids 9-32, 39-42 and 64-88 of HCV-1
XX FT and amino acids 67-84 of HCV-2"
XX FT
XX FT US2002146685-A1.
XX PN
XX XX
XX PD 10-OCT-2002.
XX XX
XX PF 14-JUN-2001; 2001US-0881654.
XX XX
XX PR 15-JUN-2000; 2000US-212082P.
XX PR 02-APR-2001; 2001US-280811P.
XX PR 02-APR-2001; 2001US-280867P.
XX XX
XX PA (CHIE//) CHIEN D Y.
XX PA (ARCA//) ARCANGEL P.
XX PA (TAND//) TANDESKE L.
XX PA (GEOR//) GEORGE-NASCIMENTO C.
XX PA (COIT//) COIT D.
XX PA (MEDI//) MEDINA-SELBY A.
XX XX

```

PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Colt D;  
PI Medina-Selby A;  
XX  
DR WPI; 2003-147573/14.  
DR N-PSDB; ABX14411.  
XX  
XX Immunassay solid support for detecting Hepatitis C Virus infection in  
PT biological samples, comprises Hepatitis C Virus conformational epitope  
PT and multiple epitope fusion antigen -  
XX  
PS Claim 25; Fig 5A-5F; 45pp; English.  
XX  
XX The present invention relates to immunoassays comprising Hepatitis C  
CC virus (HCV) NS3/4a conformational epitope and multiple epitope fusion  
CC antigen (MEHA), bound to a solid support. The NS3/4a epitope and/or  
CC the multiple epitope fusion antigen react with anti-HCV antibodies  
CC present in a biological sample from an HCV-infected individual. The  
CC immunoassays and methods of the invention are useful for detecting  
CC HCV infection in a biological sample. The inventive immunoassay solid  
CC support provides a sensitive and reliable method for detecting early  
CC HCV seroconversion. The assays can detect HCV infection caused by any  
CC six known genotypes of HCV. The use of the multiple epitope fusion  
CC proteins decreases masking problems, improves sensitivity in detecting  
CC antibodies by allowing a greater number of epitopes on a unit area  
CC of substrate, and improves selectivity. The present sequence  
CC represents HCV multiple epitope fusion antigen 7.1 (MEHA 7.1), a  
CC mutant HCV polypeptide derived from various regions of HCV type 1,  
CC 2, or 3 (HCV-1, HCV-2, or HCV-3) polypeptide sequences.  
XX  
SQ Sequence 1099 AA;  
XX  
XX Query Match 28.2%; Score 512.5; DB 24; Length 1099;  
Best Local Similarity 62.8%; Pred. No. 2,8e-35;  
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;  
XX  
QY 167 KPAIIPDREVLVYREFDEME--SOLPYIEGMMALAEQFKOKALAFASRGNHVAPTHYTE 225  
DB 701 KPAIIPDREVLVYREFDEMECSHLPYIEGMMALAEQFKOKALG--SRG----- 748  
QY 226 SDAKPLVPDKKVLVYQYDEMEAFASRGNHVAPTHYVESDASALPYMDTRALAGQF 285  
DB 749 --GKPAIVDPKEVLVYQYDEME-----CSQAPYIEQAQVIAHQF 787  
QY 286 KEKVLAFASRGNHVSRHYVPESEPOVVTTPDKEILYEAFDEME--ASKRALIEEQRMA 344  
DB 788 KEKVLGLI-----DNDQVVTTPDKEILYEAFDEMECASRALIEEQRMA 833  
345 EMLSKTIQ 352  
834 EMLSKTIQ 841  
XX  
XX RESULT 5  
ID AAR30616 standard; Protein; 3010 AA.  
XX  
AC AAR30616;  
XX  
XX 25-MAR-2003 (updated)  
DT 19-MAR-1993 (first entry)  
XX  
DE Polypeptide coded by Korean HCV full cDNA sequence LBCT.  
XX  
XX KHCV-LBCT; diagnosis; vaccine.  
XX  
XX Korean hepatitis C virus.  
XX  
XX OS  
XX PN EP52318-A2.  
XX  
XX PD 07-JUN-1993.  
XX  
XX PF 10-JUN-1992; 92EP-0109753.  
XX

PR 10-JUN-1991; 91KR-0009510.  
PR 06-AUG-1991; 91KR-0013601.  
XX  
XX (LUCK-) LUCKY LTD.  
XX  
XX PA  
XX PI Cho JM, Choi DY, Kim CH, Kim ST, Lee YB, Lim KJ, Park YM;  
PI So HS, Yang JY;  
XX  
XX WPI; 1993-001883/01.  
DR N-PSDB; AAQ33282.  
XX  
XX PT DNA and polypeptide(s) from a new type of hepatitis C virus (HCV)  
PT : for diagnosing and vaccinating against KHCV infections  
XX  
XX PS Disclosure; Fig 2; 11pp; English.  
XX  
XX The polypeptide is that encoded by the full cDNA sequence of Korean  
CC hepatitis C virus (KHCV) cDNA, KHCV-LBCT. It or its fragments may be  
CC used in a specific and accurate method for detecting KHCV antibodies  
CC in the serum of hepatitis C patients. Antibodies directed against these  
CC polypeptides are useful for the purification of KHCV antigens and for  
CC the development of an improved diagnostic to detect KHCV antigens in a  
CC sample. The polypeptides may also be used in a vaccine for treatment  
CC and prevention of KHCV infection at a dosage of 5-200 ug/peptide.  
CC (updated on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 3010 AA;  
XX  
XX Query Match 14.7%; Score 268; DB 14; Length 3010;  
Best Local Similarity 30.7%; Pred. No. 1.3e-13;  
Matches 102; Conservative 26; Mismatches 88; Indels 116; Gaps 14;  
XX  
QY 23 RPAVIPDREVLVYQYDEME--ASHLPYIEGMMALAEQFKOKALAF--ASRGNHVSPHY 78  
DB 1691 KPAIIPDREVLVYQYDEMECSHLPYIEGMMALAEQFKOKALGILQTRKQAE----- 1745  
QY 79 VPESARPAIIPDREVLVYREFDEMEAFASR--GNHVSFAHYVESDASQAAPYIEQAQVI 137  
DB 1746 ---AAP-----VVSCKPALETFFAKMMNMTISGIVLAGLSTLPNPAIRSP--- 1790  
QY 138 AHQFEKVLAFASRGNHSPSTHYVESDAKPAIIPDREVLVYREFDEMESSH--LPYIEG 196  
DB 1791 -----MAFTA-----STSTPL-----TTGHTLLFNILG 1813  
QY 197 MMLAEQFKOKALAFASRGNHVAPTHYTESDAK-----DALVPDKKVLVYQ 242  
DB 1814 GVVAAQLAPSAASAFVAGIAGAAGVITGLGKVLVDILAGYGAGVAGALVAFK--IMSGE 1872  
QY 243 YDEME-----AFPSRG-----NHAAPTHYVESDASALPYMDTRALAG 283  
DB 1873 MPSAEDVNLPLPAILSPALVGIACAILRRHVGPGGAV----- 1913  
QY 284 QFEKVLAFASRGNHVSPHYVPESEPOVVT 315  
DB 1914 QMMNRLLIAFASRGNHVSPHYVPESEPARPT 1945  
XX  
XX RESULT 6  
ID AAR82694 standard; Protein; 3010 AA.  
XX  
AC AAR82694;  
XX  
XX 14-NOV-1996 (first entry)  
DT  
XX  
XX DE Partial HCV non-structural polypeptide.  
XX  
XX KM proteinase; hepatitis C virus; screening; inhibitor; proteolytic;  
XX identification; cleavage.  
XX  
XX OS Hepatitis C Virus.  
XX  
XX Key Location/Qualifiers  
FH

```

FT      Protein      898..1233
FT      /note="partial proteinase; see AAR82692"
FT      Protein      992..1907
FT      /note="partial proteinase; see AAR82693"
XX      JP07184648-A.
XX      25-JUL-1995.
XX      PD
XX      PF
XX      05-FEB-1993; 93JP-0018854.
XX      PR
XX      04-DEC-1992; 92JP-0325303.
XX      07-FEB-1992; 92JP-0022657.
XX      18-SEP-1992; 92JP-0249240.
XX      (KAEN/) KAENNO K.
XX      (SOYA-) SOYAKU GIJUNSU KENKYUSHO KK.
XX      (SUMO) SUMITOMO METAL IND LTD.
XX      WPI, 1995-287962/38.
XX      N-PSDB; AAT03960.

PT      An HCV proteinase active substance - which has activity as an
PT      anti-HCV agent and can be used to screen for proteinase inhibitors
XX      Disclosure; Page 39-48; 52pp; Japanese.
XX      The present sequence is a partial Hepatitis C Virus (HCV) polyprotein
XX      from the non-structural region. Partial proteinase sequences (AAR82692-
XX      93) are contained within this sequence. The proteinases can be used as
XX      anti-HCV agents. They can also be used to screen cpds. for their ability
XX      to inhibit their proteolytic activity. In this way proteinase inhibitors
XX      can be identified.
XX      Sequence 3010 AA;

Query Match      14.5%; Score 263; DB 16; Length 3010;
Best Local Similarity 31.8%; Pred. No. 3.6e-13;
Matches 89; Conservative 24; Mismatches 63; Indels 104; Gaps 11;

QY      23 RPAVTPREVLYOEFDEMEE-ASHLPYIEQGMQLAEQKOKALAF--ASQ----- 70
DB      1691 RPAVTPREVLYOEFDEMEECASHLPYIEQGMQLAEQKOKALGLQTATQAEAAAFV 1750
QY      71 -----NHVSPTHYVPESDAPPAIIPREVLYHREFDEMEAFASRGNHVS 114
DB      1751 ESKRALEVFMAKMMNFISGIQYH-----AGLSTLPGNPAT-----ASLMAFTASITS 1799
QY      115 P-----AHYVPESDAS-----QAPYIEQAVIAHQFKEKVLAF 148
DB      1800 PLTTQNTLLFNILGQVAAQLAPPSAASAFVAGIAGAAVGSIGLKVLAH-----LLAG 1854
QY      149 ASRGNHDSPTHYVPESDAKPAIIPREVLYHREFDEMESQHL--PYIEQGMML----- 199
DB      1855 YGAGVAGALVAF-----KVMSEGMSTEDLVNLLPALISFGALVGVCA 1900
QY      200 -----AEQKOKALAFASRGNHVAPTHYVTESDA 228
DB      1901 ILRRHVGPGEAVQVMNRLIAFASRGNHVSPTHYVPESDA 1940

RESULT 7
AAR34468
ID      AAR34468 standard; Protein; 3011 AA.
XX      AAR34468;
XX      30-JUL-1993 (first entry)
XX      Encoded by full-length Hepatitis C virus clone JKI-B.
XX      HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
XX      polymerase chain reaction; diagnostic method.

```

```

XX      XX
OS      Hepatitis C virus.
XX      Key      location/Qualifiers
XX      Key      location/Qualifiers
XX      Misc-difference 2414
XX      /note="not defined"
XX      JP05068562-A.
XX      23-MAR-1993.
XX      PD
XX      PF
XX      30-MAY-1991; 91JP-0153736.
XX      PR
XX      30-MAY-1991; 91JP-0153736.
XX      (SANW) SANWA KAGAKU KENKYUSHO CO.
XX      WPI, 1993-130638/16.
XX      DR
XX      N-PSDB; AAQ04026.

PT      DNA and cDNA of hepatitis C virus - useful as probes for
PT      diagnosing HCV infection
XX      Claim 3; Page 6-18; 44pp; Japanese.
XX      cDNA was prepared from HCV genomic RNA. Full-length clone JKI-B
XX      and 14 shorter clones were isolated by PCR amplification.
XX      primer/probes derived from the sequences of these clones can be used
XX      in diagnostic assays for HCV. See also AAQ04025-Q40439.
XX      Sequence 3011 AA;

Query Match      14.5%; Score 263; DB 14; Length 3011;
Best Local Similarity 32.5%; Pred. No. 3.6e-13;
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 11;

QY      23 RPAVTPREVLYOEFDEMEE-ASHLPYIEQGMQLAEQKOKALAF--ASRGNHVSPTHY 78
DB      1691 RPAVTPREVLYOEFDEMEECASHLPYIEQGMQLAEQKOKALGLQTATQAEAAAFV 1745
QY      79 VPESDAPPAIIPREVLYHREFDEMEEAFASR--GNHVSPTHYVPESDASQAPYIEO-- 133
DB      1746 -----AAPVVESKQAL-----EAFVAKMMNFISGIQYHAGLSTLPGNPATVSLMA 1792
QY      134 -AQVTHQKKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB      1793 FTASITSPLTTQNTLLFNILGQVAAQLAPPSAASAFVAGIAGAAVGSIGLKVLDIL 1852
QY      171 -----IPREVLYHREFD-EMESQHL-----PYIEQGMML-----A 200
DB      1853 AGYGAGVAGALVAFKMGSEGMSTEDLVNLLPALISFGALVGVCAAILRRHVGPGEA 1912
QY      201 EQKOKALAFASRGNHVAPTHYVTESDA 228
DB      1913 VQVMNRLIAFASRGNHVSPTHYVPESDA 1940

RESULT 8
AAR68864
ID      AAR68864 standard; Protein; 3010 AA.
XX      AAR68864;
XX      06-DEC-1995 (first entry)
XX      Hepatitis C virus RNA helicase.
XX      Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;
XX      baculovirus; recombinant production.
XX      Hepatitis C virus.
XX      Key      Location/Qualifiers

```

```
FT Region 196..198
FT /label= N-linked glycosylation site
FT 209..211
FT /label= N-linked glycosylation site
FT Region 234..236
FT /label= N-linked glycosylation site
FT 250..252
FT /label= N-linked glycosylation site
FT Region 305..307
FT /label= N-linked glycosylation site
FT 325..327
FT /label= N-linked glycosylation site
FT Region 417..419
FT /label= N-linked glycosylation site
FT 423..425
FT /label= N-linked glycosylation site
FT 430..432
FT /label= N-linked glycosylation site
FT Region 448..450
FT /label= N-linked glycosylation site
FT 532..534
FT /label= N-linked glycosylation site
FT 556..558
FT /label= N-linked glycosylation site
FT Region 576..578
FT /label= N-linked glycosylation site
FT 623..625
FT /label= N-linked glycosylation site
FT 645..647
FT /label= N-linked glycosylation site
FT Region 1213..1215
FT /label= N-linked glycosylation site
FT 1255..1257
FT /label= N-linked glycosylation site
FT Region 2041..2043
FT /label= N-linked glycosylation site
FT 2077..2079
FT /label= N-linked glycosylation site
FT Region 2240..2242
FT /label= N-linked glycosylation site
FT 2788..2790
FT /label= N-linked glycosylation site
FT Region
FT JP06319583-A.
FT 22-NOV-1994.
FT 18-SEP-1992; 92JP-0249241.
FT 18-SEP-1992; 92JP-0249241.
FT (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
FT WPI; 1995-040330/06.
FT N-PSDB; AA081559.
FT Expression of hepatitis C virus helicase gene in baculovirus -
FT useful for large scale prodn. of RNA helicase.
FT Claim 1; Fig 1-4; 9p; Japanese.
FT AA081559 encodes AAR68864 hepatitis C virus (HCV) RNA helicase. The
FT DNA was used in the construction of an expression vector, which
FT was used to transform a baculovirus host. The transformed
FT baculovirus could then be used for the recombinant prodn. of
FT HCV RNA helicase.
FT Sequence 3010 AA;
FT Query Match 14.3%; Score 260; DB 16; Length 3010;
FT Best Local Similarity 32.2%; Pred. No. 6.4e-13;
FT Matches 86; Conservative 28; Mismatches 75; Indels 78; Gaps 10;
```

```
QY 23 RPAVTPDEEVLYQEFDEMEE-ASHIPTYEQGMQLAEQKOKALAF---ASRONHVSPTHY 78
DB |||||
DB 1691 RPAVTPDEEVLYQEFDEMEECASHLPTYEQGMQLAEQKOKALGLOTATKQAE----- 1745
QY 79 VPESDAPPAIIPDEEVLYQEFDEMEEAFAFR-GNHVSPAHVPESDASQAPYIQ----- 133
DB 1746 -----AAP-----VFSKRALEVFPAKHMNMNPTISGIQYLAGSTLPGNPATLSLAF 1793
QY 134 -AQVIAHOPKEKVLAF-----ASRGNHDSPTHYVESDAKPAI----- 170
DB 1794 TASITSPLTQNTLFLNIGWVAQAQAPPSAFAVGAGIAGAIVGSIQKVLVDILA 1853
QY 171 -----IPDEEVLYQEFDEMEESQHL--PTYEQGMQL-----AE 201
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLPALISPGALVVGVCMAILRRHVPGEGAV 1913
QY 202 QFKOKALAFASRGNHVAPTHYVESDA 228
DB 1914 QMMNRLLAFASRGNHVSPTHYVESDA 1940
```

```
RESULT 9
AAP90158
ID AAP90158 standard; protein; 1786 AA.
```

```
XX AAP90158;
AC 25-MAR-2003 (updated)
DT 10-NOV-1989 (first entry)
```

Protein sequence of hepatitis c virus composite cDNA.

Hepatitis C virus; vaccine.

Pan troglodytes.

GB212511-A.

26-JUL-1989.

18-NOV-1988; 88GB-0027024.

18-NOV-1987; 87US-0122714.

30-DEC-1987; 87US-0139886.

26-FEB-1988; 88US-0161072.

26-OCT-1988; 88US-0263584.

(CHIR ) CHIRON CORPORATION.

Houghton M, Choo QL, Kuo G;

WPI; 1989-215054/30.

N-PSDB; AAN90327.

Hepatitis C virus gene - used for prodn. of polynucleotide probes,

polypeptide(s) and antibodies for diagnosis, prevention and treatment

of infection.

Disclosure; fig 26; 30pp; English.

The sequence is encoded by the composite cDNA of AAN90327. These

antigens react with antibodies in patients with non-A non-B hepatitis

(NANBH). They can be used to diagnose HCV-induced NANBH, to raise

antibodies for immunoassay or treatment, or to produce vaccines.

(Updated on 25-MAR-2003 to correct PR field.)

Sequence 1786 AA;

Query Match 14.2%; Score 259; DB 10; Length 1786;

Best Local Similarity 31.5%; Pred. No. 3.8e-13;

Matches 84; Conservative 28; Mismatches 77; Indels 78; Gaps 9;

QY 23 RPAVTPDEEVLYQEFDEMEAS-HLPTYEQGMQLAEQKOKALAF---ASRONHVSPTHY 78

Query Match	14.2%	Score 259;	DB 14;	Length 2354;
Best Local Similarity	32.2%	Pred. No. 5.6e-13;		
Matches	86;	Conservative	30;	Mismatches 73; Indels 78; Gaps 10.
OY	23	RPAAVTPREVLYOEFDEMEE-ASHLPYIEQGMOLAEQFKOKALAF----	ASRQNHVSPTHY	78
		:		
DB	875	RPAAVTPREVLYOEFDEMEECASHLPYIEQGMOLAEQFKOKALGLLOTATKQAEA-----		929
OY	79	VPESDARPAIIPREVLRHREFDEMEEAFASR-GNHVSPAHYVP-----	ES	122
		:		
DB	930	----AAP-----VVESKKRALETFTWAKIMNFIISGIQYLAGLSTLPGNPAIASLMAF		977
OY	123	DASQAPAYIEQAYI-----	AHQFEKYLAFASRGNHDSPTHYVPESD	165
		: :	:	
DB	978	TASVTSPLTQSTLLNLILGQWVAOLAPSAASAFVQAGIAGAAGVSGIGKVLVDILA		1037
OY	166	AKPALIIDREVLVYEFDP-EMEESSH-----PYIEQGMML-----	AE	201
		: :	:	
DB	1038	GYGAGVAGALVAFKVMSEGPSTEDLVNLLPALISPGALVGVCAALIRRHVGPBEGAV		1097
OY	202	QFKOKALAFASRGNHVAPTHYVTESDA		228
		: :	:	
DB	1098	QMMNRLLAFASRGNHVSPTHYVAESDA		1124
RESULT 11				
AAR29527				
ID	AAR29527	standard; Protein: 2510 AA.		
XX	AAR29527;			
AC				
XX				
DT	25-MAR-2003	(updated)		
DT	26-APR-1993	(first entry)		
XX				
DE	HCV antigen T7N1-30.			
XX				
KN	Clone; Hepatitis C Virus; HCV, core-envelope; NS1(gp70); NS2-NS4;			
KW	NS4-NS5; region; diagnostic method; antibody; suppress; control;			
KW	proteolytic; process; precursor; polypeptide.			
XX				
OS	Hepatitis C virus.			
XX				
FH	Key	Location/Qualifiers		
FT	Misc-difference 2212			
FT	/note= "Nonsense codon"			
XX				
PN	EP518313-A2.			
XX				
PD	16-DEC-1992.			
XX				
PF	11-JUN-1992;	92EP-0109812.		
XX				
PR	11-JUN-1991;	91JP-0139268.		
PR	12-JUL-1991;	91JP-0172794.		
PR	07-OCT-1991;	91JP-0287008.		
PR	16-DEC-1991;	91JP-0332329.		
PR	20-APR-1992;	92JP-0099957.		
XX				
PA	(MTTU ) MITSUBISHI KASEI CORP.			
XX				
PI	Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;			
PI	Teranishi Y;			
XX				
DR	WPI; 1992-417213/51.			
XX				
DR	N-PSDB; AAQ32436.			
XX				
PT	New hepatitis C virus gene and its encoded protein - used for			
PT	diagnosing and vaccinating against hepatitis C virus infections			
XX				
PS	Claim 1 and 3; Page 259-272; 305pp; English.			
XX				
CC	This sequence was encoded by the Hepatitis C Virus (HCV) gene of the			

Sequence 2510 AA;

every Match	14.2%;	Score 259;	DB 13;	Length 2510;
Best Local Similarity	30.3%;	Pred. No. 6.1e-13;		
Matches 96;	Conservative 26;	Mismatches 109;	Indels 86;	Gaps 11;

```

23 RPAVTPDREVLYQEFDEMEE-ASHLPYIEQGMLAEQFKKALAF---ASRQNHSPTHY 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1691 RPAVIPDREVL YQEFDEMEECASHLPYIEQGM LAEQFKKALGLTATTKDAE----- 1745

```

```

79 VPESDARPAII PDREVLHREFEMEEAFASR-GNHVSPAHYVPESDASQAADYIEQAQVI 137
      | | : : | : | : |
1746 -----VESKRALETFWAKHMNFISGIQYLAGLSTLPGNPAI----- 1787
      AAP-----

```

```

138 AHQKEXVAFASRGHNHDSPTHYVPESDAKPAIIPDREVLYRFDEMEESQLPYIEGGM 197
      ::||:      ||      ::||:      ||
1788 -----ASLMAFTASITSPLTTOYT-----LTFNIGGWAAQIAPPSAASA 1828

```

198 MLEQFKKALAFASRGNHVAPTHYVESDAKPALVPDKEVLVQQYDEME-----AFA 251  
          |       |       |       |||      |      |  
1829 FVGAGITGAAVGSGIGKLVLDILAGYGAGVAGALVAFK-VMSGMPSTEDLVNLLPAIL 1867

252 SAG-----NHVAPHYVVESDASASLPYMDETRAISOQFKEXKLAFAASGNH 298  
 1888 SPGALVGVCAALIRHVEGEGAV-----QMMNRILIAFASRGNH 1928

229 VSNHIVESEFQVVI 315  
||| ||||: ||  
1929 VSPTHYVPESDAAARVT 1945

T 12  
995  
AAR66995 standard; Protein; 3011 AA.

AAR66995;  
01-AUG-1995 (first entry)

Hepatitis C virus gene HC-J1/protein; specific antibodies

Hepatitis c virus.  
JP06284887-A.

11-OCT-1994.  
10-DEC-1993; 93JP-0345753

10-DEC-1992; 92JP-0360705  
(IMMO ) IMMUNO JAPAN KK.

MP1; 1994-362594/45.  
N-PSDB; AAQ74770.

Claim 11; Pages 18-32; 35pp; Japanese.

AA074770 encodes AAR66995 the HC-J1/protein, the cDNA can be used in

the construction of an expression vector for the transformation of a host cell. The host cell can then be used in the production of proteins and peptides, useful in the preparation of monoclonal and polyclonal HCV-specific antibodies.

ry Match	14.2%;	Score 259;	DB 15;	Length 3011;
at Local	Similarity	31.7%;	Pred. No. 7.9e-13;	
ches 85;	Conservative	26;	Mismatches 75;	Indels 80;
				Gaps 10;

**23** RRAVLPDEVLQGFDEMEEAS-HLPIYEQGQLAEQFQKALAF---ASRQHNV-SEPH 77  
|||:|||::|||||::|||||::|||||::|||||::|||||::|||||::|||||:  
**1691** RBAILPDEVLYREFDEMEESQHLPYEQGMMLAEGFRQKALGLLIQTASTAQEAIVPT- 1749

```

78 YPESDARPAIPREVLFHEFDEMEAFASR-GNHVSPAHYVPESDASQAAPYIEQ--- 133
      : : : : :
1750 -----VQTNWQKLEAFWAKHMWNFISGIQYLGLSTLPGNPATASLMA 1792

```

```

134  --AQVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDARPAI----- 170
      | | : | | : | | : | | :
1793 FTAAVTSPLTTSQTLLEINIGWVAQQLAAPGATAFVGSGLAGNAVGSVGLGRVLVDIL 1852

```

```

1853 AGYGAGVAGALVAFKIMSGELPSTEDLVNLLPAILSPGALVGVCAILIRHVGPEGGA 1912
      :  ::  |  :  :  |  :  :
1854 -----IPREVLVREFFDEMBSQHL--PYIEQGMML-----A 200
1855 1/1 -----

```

201 EQFQONALAFASRGNHVAPIHYVESDA 228  
| : : ||||| | : ||| |  
1913 VQMMNRLIAFASRGNHVSPTHYVESDA 1940

AAAR12599 standard: Protein: 1250 AA  
599 T 13

AAAR12599;  
225-MAR-2003 (updated)

Portion of PT-NANBH viral non-structural protein

Non-A, non-B hepatitis virus.

226-JUN-1991.

003-MAR-1990; 90GB-0004814.  
18-DEC-1989; 89GB-0028562.  
END 1000 0000

17-DEC-1990; 90GB-0027250.  
(WELL ) WELLCOME FOUND LTD.  
RICH/A HICHTFIELD D D

Highfield PE, Rodgers BC, Tedder RS, Barbara JAJ  
 PT: 1991-197584/25

post-transfusional non-A non-B hepatitis peptide(s) - and  
SSO DNA and antibodies used in diagnostic assays and in neutralization

Claim 1; Page 88-97; 108pp; English.

The sequence was deduced from a non-structural (3') coding region sequence isolated from serum of humans infected by the PT-NANBH





This sequence is encoded by the genome of a blood transmissible non-A, non-B hepatitis (NANBH) virus. The cDNA sequence was isolated using the primers given in AA063500-35. The amplified fragments are used in the detection of hepatitis virus. The target DNA was isolated from serum of chronically infected NANBH patients who were C100 antibody-positive and HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR CC were performed on cDNA and the total human NANBH DNA was constructed from 23 clones.

Sequence 3010 AA;

Query Match 14.1%; Score 257; DB 15; Length 3010;  
Best Local Similarity 31.5%; Pred. No. 1.2e-12;  
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;

QY 23 RPAVTPDREVLYQEFDEMEE-ASHLPYIEQGMQLAEQFKOKALAF---ASRQNHVSPTHY 78  
DB 1691 RPAVTPDREVLYQEFDEMEECASHLPYIEQGMQLAEQFKOKALGLQATKQAE----- 1742  
79 VPESDARPAIIPDREVLYHREFDEMEEFASR-GNHVSPAHYVPESDASQAAPYIEQ----- 133  
1743 ---AEAAPVVESE-----WRALFAFWAKHMMNFIISGIOYLAGLSTLPGNPAIASLMAF 1793  
QY 134 -AQVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170  
DB 1794 TASITSPLTONTLLFNILGWVAAQLAPPSAAXFVGAGIAGAAIGSIGLKVLDILA 1853  
QY 171 -----IPDREVLYREFDEMEESQHL--PYIEQGMML-----AE 201  
DB 1854 GYGAGVAGALVAFKVMGSEAPSAEDLVNLLPALISPGALVXXVCAAILRRHVGPGEAV 1913  
QY 202 QFKOKALAFASRGNHVAPTHYVYESDA 228  
DB 1914 QMMNRLIAPASRGNHVSPTHYVPESDA 1940

RESULT 15

AA:68622  
ID AAR68622 standard; Protein: 3010 AA.

AC AAR68622;  
DT 16-OCT-1995 (first entry)  
XX HCV protein cleavable with new serine proteinase.  
DE HCV proteinase; serine; cleavage; hepatitis C virus; HCV.  
XX  
KM Hepatitis C Virus.

FH Key Location/Qualifiers  
FT Cleavage-site 2419..2420  
FT /note="Serine protease cleavage site"

PN JP06315377-A.

PD 15-NOV-1994.

XX 06-MAY-1993; 93JP-0105666.

PR 06-MAY-1993; 93JP-0105666.

XX (KAEN/) KAENNO K.

PA (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.

PA (SUMO) SUMITOMO METAL IND LTD.

XX WPI; 1995-032330/05.

DR N-PSDB; AA080498.

XX New HCV-originated proteinase active substance - used for  
PT site-specific cleavage by an intermolecular reaction and the  
PT purification thereof

PS Disclosure; Page 10-19; 23pp; Japanese.

XX This protein from HCV (hepatitis C virus) (encoded by AA080498) is  
CC cleaved between amino acids 2419 and 2420, by a new serine protease,  
CC contg. the sequence of AAR68621. The proteinase is purified as a fused  
CC product with the dihydrofolate reductase protein by using a methotrexate  
CC column. It can be used for the development of an inhibitor for HCV  
CC proteinase.  
XX

Sequence 3010 AA;

Query Match 14.1%; Score 257; DB 16; Length 3010;  
Best Local Similarity 31.8%; Pred. No. 1.2e-12;  
Matches 85; Conservative 29; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPAVTPDREVLYQEFDEMEE-ASHLPYIEQGMQLAEQFKOKALAF---ASRQNHVSPTHY 78  
DB 1691 RPAVTPDREVLYQEFDEMEECASHLPYIEQGMQLAEQFKOKALGLQATKQAE----- 1745  
79 VPESDARPAIIPDREVLYHREFDEMEEFASR-GNHVSPAHYVPESDASQAAPYIEQ----- 133  
1746 ---AAP-----VESKWRALBFVFAKHMNFIISGIOYLAGLSTLPGNPAIASLMAF 1793  
QY 134 -AQVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170  
DB 1794 TASITSPLTONTLLFNILGWVAAQLAPPSAAXFVGAGIAGAAVGSIGLKVLDILA 1853  
QY 171 -----IPDREVLYREFDEMEESQHL--PYIEQGMML-----AE 201  
DB 1854 GYGAGVAGALVAFKVMGSEAPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGEAV 1913  
QY 202 QFKOKALAFASRGNHVAPTHYVYESDA 228  
DB 1914 QMMNRLIAPASRGNHVSPTHYVPESDA 1940

Search completed: July 25, 2003, 13:09:44  
Job time : 87 secs